Blind Deconvolution With An Application To Speckle Imaging

by

Bonnie Ngai Fong Law, B.E.(Hons.)

Department of Electrical and Electronic Engineering

Submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

University of Tasmania

September 1996

Statement of Originality

This thesis contains no material which has been accepted for the award of any other degree or diploma in any tertiary institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except when due reference is made in the text.

Bonnie Ngai Fong Law

Authority of Access

This thesis may be made available for loan and limited copying in accordance with the Copyright Act 1968.

Bonnie Ngai Fong Law

Abstract

It is well known that the quality of an image that can be obtained from a ground-based telescope is limited by the wavefront distortion associated with the atmospheric turbulence. The turbulence is present as a result of the mixing of the warm and the cold air. Although the turbulence affects both the amplitude and the phase of the wavefront, the random phase aberration is found to be more important in the distortion of an image than the magnitude distortion. As a result, the available light spread over a wide area and the observed image is blurred.

This effect has hindered astronomers since the invention of the telescope in the 17^{th} century. In order to reduce the effect of the atmospheric turbulence and improve the quality of the image obtained at the ground-based telescopes, large collection aperture telescopes have been built at high altitude places. Although the light gathering properties of these modern telescopes are remarkable, the problem that originated from the atmospheric turbulence has not been solved completely.

The aim of this study is to find ways that can compensate the effects of atmospheric turbulence and provide a good quality image. There are two approaches proposed to deal with this problem, namely the adaptive optics and the post-detection processing approaches. The first half of the thesis is concerned with an adaptive optics approach while the second half is concerned with a post-detection processing approach.

The adaptive optics approach compensates the wavefront distortion in real time. It uses hardware designed to physically alter the optical path length of the distorted wavefront. Thus a disturbance which is in the optical path just opposite to the disturbance associated with the atmospheric turbulence is introduced so that both disturbances cancel. Therefore, an accurate estimation of the disturbance introduced by the atmospheric

turbulence is crucial in this approach. While the problem can be reformulated as a standard least square problem from the measurements from Shack Hartmann sensor, the least square problem is itself under-determined. This means that the quality of distortion is limited by the number of measurements available, which in turn is limited by the available light.

In order to solve this under-determined problem, a priori information needs to be applied so as to put more constraints into the estimation. One piece of a priori information that can be obtained is the statistics of the turbulence. This can be derived theoretically using the Kolmogorov turbulence model. By incorporating this information into the estimation, the problem becomes well-posed. As the simulation results show, the accuracy in estimating the distorted wavefront is improved by incorporating the a priori information.

The second approach to compensate the disturbance is the post-detection processing approach. This is founded upon the insight that short exposure images preserve frequencies up to the diffraction limit. A variety of techniques have been proposed to recover this diffraction-limited information, for example, the correlation-based techniques of Knox-Thompson and bispectrum approaches. However, these techniques are required to average the effects of atmospheric turbulence before the algorithm can be applied to recover the object of interest.

Another approach is to consider the problem as a deconvolution problem in which the object of interest is reconstructed based on the disturbance information and the short exposure image. In this case the disturbance is unknown. It is a blind deconvolution problem, that means we need to reconstruct the object from the short exposure image without a statistical ensemble of distorted images. Three methods are discussed and compared, they are the conjugate gradient approach, the projection-based approach and the maximum likelihood approach. All these three methods can be formulated using the Bayes' theorem but with different assumptions made on the noise statistics and with different application of the a priori knowledge. They are compared in terms of the convergence rate, the quality of the reconstructed image and also in the presence of different noise levels.

As simulation results show, all these three algorithms can blindly deconvolve a blurred image. In general, the projection-based algorithm and the maximum likelihood algo-

rithm produce clean reconstructed images compared to the conjugate gradient algorithm. However, all three algorithms experience the so-called superresolution effect in the presence of noise. A way to prevent suprereolving a blurred image is suggested.

Acknowledgements

Firstly, I would like to thank my previous supervisor Doctor Richard G. Lane and my current supervisor Professor D. Thong Nguyen. In particular I am grateful to Doctor Lane for his guidance and contributions that have helped me greatly in my research, and also for his encouragement and support throughout these studies.

I am also very grateful to Professor D. Thong Nguyen for his guidance and encouragement. His valuable suggestions together with the experience he has in supervising students have greatly benefited me both academically and personally.

I like to thank all staff of the Department of Electrical and Electronic Engineering at the University of Tasmania, especially Doctor Habib Talhami, Mr Gregory The, Mr Peter Watt and Doctor David Lewis for their kind treatment and attending my seminars throughout the course of studies. Many thanks go to all my fellow PhD students for their help, particularly, Andrew Bainbridge-Smith, Andrew Innes and Marc Stoksik.

I am very grateful to Jeoff G. Whitting and Heather Whitting for their encouragement and friendship which make the stay in Tasmania both enjoyable and memorable. I like also to thank Ron Reynolds and Doreen Reynolds for showing me kindness and friendship during the stay in Christchurch, New Zealand. I greatly appreciate that.

I greatly appreciate the love, help, support and patience of Alan Liew. I am greatly indebted to my parents, my brother and sister for their love and encouragement given to me.

Finally, I like to extend my appreciation and thanks to the financial support from the University of Tasmania and also from the Department of Electrical and Electronic Engineering.

Glossary

This thesis contains a certain amount of mathematics. The following notations and symbols are employed.

$f(x)$,$f(x,y)$	One-dimensional signal and two-dimensional image.
$F(u)$,$F(u,v)$	Fourier transform of one-dimensional signal and two-dimensional image.
σ^2	variance.
$Pr(x)$	Probability of x.
$\longleftrightarrow$	Fourier transform pair.
$\|\|x\|\|$	Norm of x.
(x,y)	Inner product of x and y.
$< S_m(x) >$	Ensemble average of $S_m(x)$, ie., $< S_m(x) >= \frac{1}{M}\sum_{m=1}^{M} S_m(x)$
$\delta_{i,j}$	The Kronecker delta, equal 1 if $i = j$ and 0 otherwise.
$\propto$	Proportional to.
∞	Infinity.
$\in$	A member of.
$\cup$	Union of sets.
$\cap$	Intersection of sets.
Σ	Summation.
Π	Product.
$\mathbf{A}^{-1}$	Inverse of the matrix A.
$\mathbf{A}^{T}$	Matrix transpose of A.
$\odot$	Convolution operator.
$\otimes$	Correlation operator.

π	Pi (3.1415...).
j	$\sqrt{-1}$.
r_0	Fried's parameter.
t_s	Time internal when atmosphere can be considered to be frozen.
t_r	Redistribution time of atmosphere.
θ	Isoplanatic patch.
x^*	Complex conjugate of x.

The following abbreviations have also been used:

FT	Fourier Transform.
AT	Another Term.
DFT	Discrete Fourier Transform.
FFT	Fast Fourier Transform.
mse	Mean Square Error.
SNR	Signal-to-Noise Ratio.
GEM	Generalized Expectation and Maximization Algorithm.
EM	Expectation and Maximization Algorithm.
CD	Complete Data.
ID	Incomplete Data.
MAP	Maximum A Posteriori.
ML	Maximum Likelihood.
CG	Conjugate Gradient.
BD	Blind Deconvolution.
KT	Knox-Thompson Algorithm.
TC	Triple Correlation.
POCS	Projection Onto Convex Sets.
psf	Point Spread Function.
RHS	Right Hand Side.

Preface

The universe has long been a mystery to humankind. Since the telescope was invented in the 17th century, more of the universe became visible. However, the curiosity of humankind motivated the advancement of science and technology to observe the faintest and the farthest star. Despite all the efforts that have been made, the problems that originate from the transmission of the light through the turbulent atmosphere have not been overcome. Putting telescopes in space above the disturbing atmosphere, like the Hubble space telescope, would seem to be an ideal solution to this problem.

Although distant stars and galaxies can be observed clearly by Hubble space telescope, the cost is extremely high, and there is currently only one telescope. Many researchers around the world are trying to explore a way to compensate the effect of atmospheric turbulence and thus provide a picture that has a comparable quality to that obtained from the Hubble space telescope. That is the ultimate aim of this research.

Many different techniques have been proposed to recover the lost resolution. There are generally two approaches to do so. The first one is a real-time approach, generally known as adaptive optics approach. It is itself a complex technology but compared to space technologies, it costs much less. It achieves the resolution by real-time compensation using hardware, that is, by introducing disturbance which is calculated to be just the opposite of the atmospheric effects. Consequently, both disturbances cancel each other out and thus produce a good picture of the universe.

The estimation of the disturbance introduced by the atmospheric turbulence can be reformulated as a least square problem, although the problem is itself ill-posed. Following a suggestion from Dr Lane, the incorporation of a priori information about the atmospheric turbulence into the estimation was developed so that the problem became

well-posed. This developed into an algorithm to estimate the distorted wavefront and reported in (Law and Lane 1996).

The second approach to recover the lost resolution is a software approach. Here the image obtained is postprocessed by an earth-based telescope on a computer. The process is regarded as a blind deconvolution problem in which the disturbance is assumed to be unknown and both the disturbance and the object of interest are estimated simultaneously.

A few different techniques have been proposed to deal with this blind deconvolution problem, such as the conjugate gradient approach, the projection based approach and the maximum likelihood method. The comparative results of these algorithms have seldom been reported in the literature although this information is valuable when one is looking for some methods to solve a particular blind deconvolution problem. Dr Lane suggested me to compare these algorithms and the results are reported in (Law and Lane, 1996). Also, in June 1995, Prof Nguyen noticed the similarity of an idea used in the projection-based blind deconvolution to the one used in the neural networks. The consequence of using the momentum term to speed up the convergence rate of the projection-based blind deconvolution algorithm are reported in (Law and Nguyen, 1995). All these results are summarized in Chapter 6 and Chapter 7.

Thesis Organization

This thesis is organized into eight chapters. Chapter 1 gives a brief summary of the mathematical techniques that are used throughout the thesis. This chapter also demonstrates that phase is more important than the magnitude information in the recognition of images.

Chapter 2 provides a detailed introduction to the problem of astronomical imaging. In particular, the statistics of the turbulence is given, the effect of the atmospheric turbulence on the image formation of a celestial object is described and a mathematical model to describe this effect is given.

Chapter 3 reviews various techniques proposed to deal with the problem of the disturbance introduced by the atmospheric turbulence. Two general approaches to deal with

this problem, namely the real time adaptive optics approach and the post-detection processing approach are summarized.

The estimation of the disturbance introduced by the atmospheric turbulence is reformulated as a least square problem as shown in Chapter 4. The pseudo-inverse solution is presented first. This is then followed by an optimal method which uses a priori knowledge about the statistics of the atmosphere to estimate the distortion introduced by the atmospheric turbulence. This optimal method is also compared with some existing estimation methods for the adaptive optics approach.

Chapter 5 outlines some practical limitations in correcting the wavefront distortion estimated by using the optimal method discussed in Chapter 4. It also suggests a method to obtain the a priori information about the statistics of the atmosphere in the practical situation.

Chapter 6 presents three methods to blindly deconvolve a blurred image. They are the conjugate gradient method, the projection-based method and the maximum likelihood method. Two ways to improve the convergence rate of the projection-based method, namely the use of a momentum term and the use of an adaptive step size, are also discussed.

Chapter 7 presents the simulation results for the three blind deconvolution methods outlined in Chapter 6. It first compares the convergence rate of the original projection-based method with the accelerated algorithm using the momentum term and the adaptive step size. Then the three blind deconvolution methods are compared in terms of the convergence rate and the quality of the reconstructionts in the presence of different noise levels.

Finally, Chapter 8 contains a summary of the major results and suggestions for future research.

Supporting Publications

This research resulted in a number of journal and conference publications during the course of this study. These are listed below:

1. N.F. Law and R.G. Lane, "Wavefront Estimation at Low Light Levels", Optics Communication, Vol. 126, pp. 19-24, 1996.

2. N.F. Law and R.G. Lane, "Blind Deconvolution Using Least Squares Minimisation", Optics Communication, Vol. 128, pp. 341-352, 1996.

3. N.F. Law, A. Liew and D.T. Nguyen, "Multiresolution Deconvolution Using Conjugate Gradient", to be published in IEEE Singapore International Conference on Communication Systems 1996.

4. N.F. Law and D.T. Nguyen, "Multiple Frame Projection Based Blind Deconvolution", in *Electronics Letter*, Vol. 31, No. 20, pp. 1734–1735, 28th September, 1995.

5. N.F. Law and D.T. Nguyen, "Improved Convergence of the Projection Based Blind Deconvolution", in *Electronics Letter*, Vol. 31, No. 20, pp. 1732-1733, 28th September, 1995.

6. N.F. Law and D.T. Nguyen, "Accelerated Multi-Frame Projection Based Blind Deconvolution", in *Proceedings of DICTA-95* (Anthony Maeder and Brian Lovell eds.), pp. 174-179 , Australian Pattern Recognition Society, Dec., 1995.

7. R.G. Lane, N.F. Law and A. Bainbridge-Smith, "Ensemble Deconvolution using a Wavefront Sensor", in *Proceedings of DICTA-93* (K. K. Fung and A. Ginige, eds.), pp. 236–243, Australian Pattern Recognition Society, Dec., 1993.

Contents

List of Figures

List of Tables

Chapter 1

Mathematical Preliminaries

Fourier analysis is a mathematical tool of great utility in the analysis of both linear and nonlinear phenomena. It has been widely used in the study of electrical networks, communication systems and many other physical systems. By using such a transform, many operations, such as convolution and correlation, that are difficult to carry out directly in the original space can be simplified in the transformed space. Mathematically, the Fourier transform can be considered either as a method of changing the co-ordinate system or equivalently a one-to-one mapping from one space to the other space in which the energy is preserved through the transformation. A brief introduction to the Fourier transform is given in Section 1.1.

The Fourier transform also provides a model of the diffraction effects which occur when an object is imaged through a finite size aperture. The diffraction phenomena can be divided into two general classes, depending on the distance of the object to the aperture,

1. Fraunhofer diffraction which occurs when the source of light is at a large distance from the observing aperture.
2. Fresnel diffraction which occurs when the source of light is at a lesser distance from the aperture, aperture.

The Fresnel diffraction is a parabolic approximation to a wavefront while the Fraunhofer diffraction is a planar approximation. However, both Fresnel and Fraunhofer diffraction

patterns can be described through the Fourier transform operation as is shown in Section 1.2.

Because of the widespread use of circular apertures, telescopes, and lenses, there is a need to model wavefront distortions using functions defined on a circular range. Zernike polynomials are a set of polynomials defined on a unit circle and they are separable in terms of angle and radius. One of the important properties of these polynomials is that they form an orthonormal basis function over a circular aperture, and thus is a popular choice to model wavefront distortions. Section 1.3 gives a brief introduction to Zernike polynomials.

Since many complex decision problems can be analysed through transforming these problems onto optimization problems, one can approach the problem of estimating the wavefront distortion in the same way. Using the optimization philosophy, an objective function is designed to quantify the performance of removing the wavefront distortion. These inter-related unknowns are thus estimated as an objective function to be maximized, or minimized depending on the formulation of the problem, subject to constraints that may limit the selection of these inter-related values. The constraints could be any measurements or observations available that are related to the unknowns.

There are many techniques which have been developed to solve this kinds of problems. One common way to find the unknowns is a maximum likelihood approach. It is a statistical approach to the problem. It tries to find a solution that not only is consistent with the observations or measurements, but also has the highest probability of causing those observations or measurements. The objective function to be maximised in this case is called the likelihood function that needs to be maximized as a function of the appropriate unknowns. This approach is discussed in Section 1.4.1.

Another approach to find the unknowns is commonly known as the descent method. It searches for those unknowns that are in the turning point of the objective function. There are two classical approaches to perform the searching, they are the steepest descent search and the conjugate gradient search. Section 1.4.2 covers both of them.

Another approach to the optimization problem, which is discussed in Section 1.4.3, is the projection method. It reformulates the problem so that all requirements or constraints that the unknowns have to satisfy are represented as mathematical sets. And then it looks for a solution that lies in the intersection of these sets. It in turn

means that the solution obtained attempts to satisfy all requirements simultaneously.

1.1 Fourier Transform

In this thesis, an image and its Fourier transform are described by complex scalar functions, denoted by lower-case and upper-case characters such as $f(\mathbf{x})$ and $F(\mathbf{u})$, where $\mathbf{x}$ and $\mathbf{u}$ are position vectors locating arbitrary points in the multi-dimensional image space and Fourier space respectively. The symbol, $\longleftrightarrow$, is used to denote a multi-dimensional Fourier transform relation, ie.,

$$F(\mathbf{u}) \longleftrightarrow f(\mathbf{x}) \tag{1.1}$$

The Fourier transform is defined as,

$$F(\mathbf{u}) = \int(K) \int f(\mathbf{x}) \exp(-j2\pi \mathbf{u} \cdot \mathbf{x}) d\sigma(\mathbf{x}) \tag{1.2}$$

where $\int(K)\int$ denotes a K-dimensional integral, $d\sigma(\mathbf{x})$ is the volume element in K-dimensional image space, and $\mathbf{u} \cdot \mathbf{x}$ denotes the dot product between vectors $\mathbf{u}$ and $\mathbf{x}$, ie.,

$$\mathbf{u} \cdot \mathbf{x} = \sum_i \mathbf{u}_i \mathbf{x}_i \tag{1.3}$$

The inverse Fourier transform is defined by,

$$f(\mathbf{x}) = \int(K) \int F(\mathbf{u}) \exp(j2\pi \mathbf{u} \cdot \mathbf{x}) d\sigma(\mathbf{u}) \tag{1.4}$$

Convolution plays an important role in both signal and image processing. It relates the output of a linear, time-invariant system to the input of the system and its point spread function (psf) [1]. The convolution relationship can be expressed in two ways as,

$$\begin{aligned} f(\mathbf{x}) &= f_1(\mathbf{x}) \odot f_2(\mathbf{x}) \\ &= \int(K) \int f_1(\mathbf{s}) f_2(\mathbf{x} - \mathbf{s}) d\sigma(\mathbf{s}) \\ &= \int(K) \int f_2(\mathbf{s}) f_1(\mathbf{x} - \mathbf{s}) d\sigma(\mathbf{s}) \end{aligned} \tag{1.5}$$

where the symbol $\odot$ denotes convolution, $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are hereafter referred to as components of the convolution $f(\mathbf{x})$. If both sides of Eq (1.5) are Fourier transformed, it can be shown that,

$$F(\mathbf{u}) = F_1(\mathbf{u})F_2(\mathbf{u}) \tag{1.6}$$

ie.,

$$f_1(\mathbf{x}) \odot f_2(\mathbf{x}) \longleftrightarrow F_1(\mathbf{u})F_2(\mathbf{u}) \tag{1.7}$$

This transform pair relationship is particularly useful in signal and image processing. Carrying out convolution in image space would be time-consuming and ineffective especially for large images. But because the convolution of two functions in image space is equivalent to the multiplication in Fourier space, it is usually easier and faster to do the convolution by multiplying the Fourier transforms of the two images and then applying the inverse Fourier transform to obtain the convolved result.

The autocorrelation function of $f(\mathbf{x})$ is a special form of convolution. It is defined by [2],

$$\begin{aligned} ff(\mathbf{x}) &= f_1(\mathbf{x}) \otimes f_1(\mathbf{x}) \\ &= f_1(\mathbf{x}) \odot f_1(-\mathbf{x}) \\ &= \int (K) \int f(\mathbf{s}) f(\mathbf{x}+\mathbf{s}) d\sigma(\mathbf{s}) \end{aligned} \tag{1.8}$$

where the symbol $\otimes$ is used to denote the correlation operator. The autocorrelation theorem [2] states that the autocorrelation function of $f_1(\mathbf{x})$, and the modulus of the Fourier transform of the function, $|F_1(\mathbf{u})|^2$, form a Fourier transform pair, ie.,

$$|F_1(\mathbf{u})|^2 \longleftrightarrow f_1(\mathbf{x}) \otimes f_1(\mathbf{x}) \tag{1.9}$$

If one wants to reconstruct an image $f_1(\mathbf{x})$, both magnitude and phase of $F_1(\mathbf{u})$ are usually required. As can be seen from Eq (1.9), the information contained in the square magnitude of the Fourier transform of $f_1(\mathbf{x})$ is same as that contained in the autocorrelation function. Since the magnitude of a function must be real, it contains no phase information in the Fourier space. This implies that the phase information about $F_1(\mathbf{u})$ is also completely lost in its autocorrelation function.

The importance of the phase information can be demonstrated by the following example. Consider an image shown in Figure 1.1. The Fourier transform of the image is

shown in Figure 1.2. If the spectral magnitude of the Fourier transform of the image is replaced by a constant and the phase is kept unchanged, the resultant image is shown in Figure 1.3. If now the phase of the Fourier transform of the image is replaced with some random numbers but the spectral magnitude is kept unchanged, the resultant image is shown in Figure 1.4. Comparing the images shown in Figure 1.1, Figure

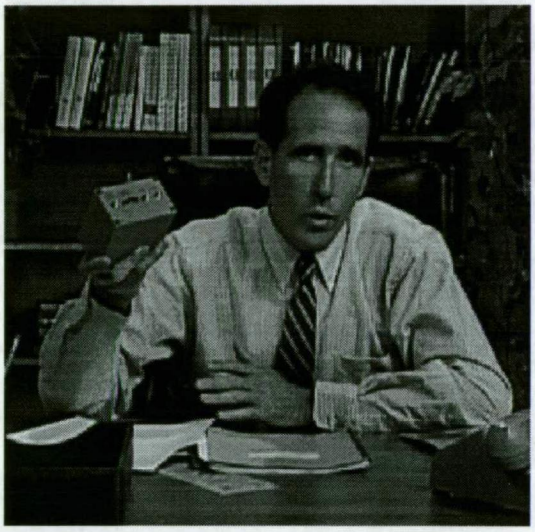
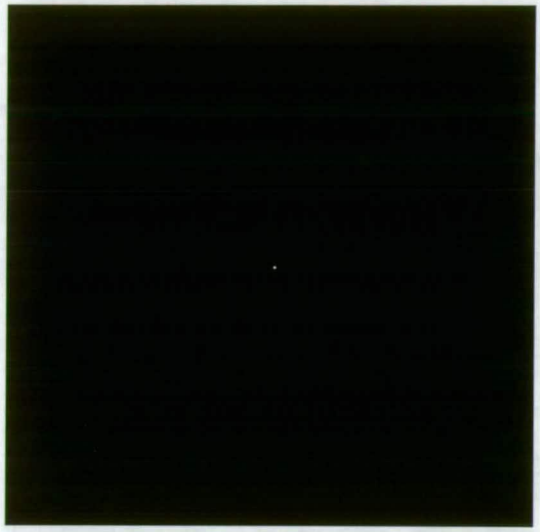


Figure 1.1: The image used to demonstrate the importance of the phase information in image recognition.

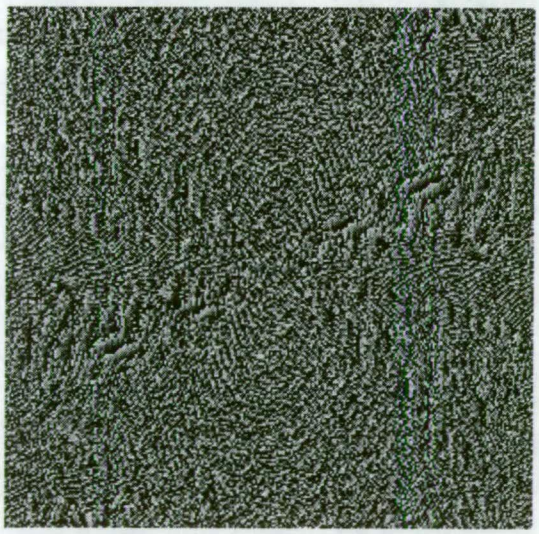
1.3 and Figure 1.4, the image that retains the Fourier phase can still be recognised. However, the shape of the image that retains the Fourier magnitude changes and is significantly different from the original one. Phase information is thus more important than the magnitude information in recognizing and also reconstructing an image [3]. In particular, it is difficult to reconstruct an image from the autocorrelation data alone as no phase information is contained in the autocorrelation function. If one wants to reconstruct the image $f_1(\mathbf{x})$ from its autocorrection function, then one has to employ methods so that the phase information is recovered. This problem is commonly known as phase retrieval and is covered in Section 3.4.

1.2 Fresnel And Fraunhofer Diffraction Patterns

The Fourier transform is found to be closely related to the Fresnel and Fraunhofer diffraction patterns. This can be shown by consideration of the diffraction geometry



(a)



(b)

Figure 1.2: (a) The magnitude and (b) the phase of the Fourier transform of Figure 1.1.

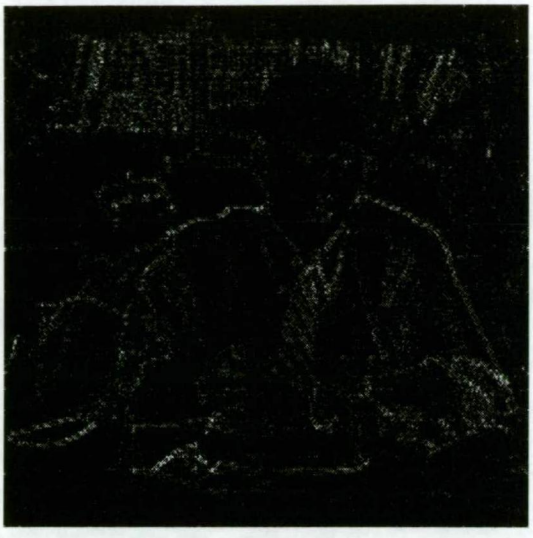


Figure 1.3: The reconstructed image formed by replacing all the spectral magnitude to 1 but keeping the phase information unchanged.

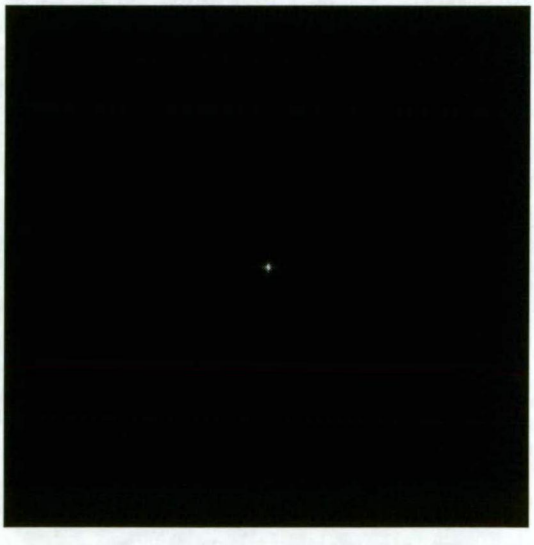


Figure 1.4: The reconstructed image formed by replacing all the phases to a set of random numbers but keeping the magnitude information unchanged.

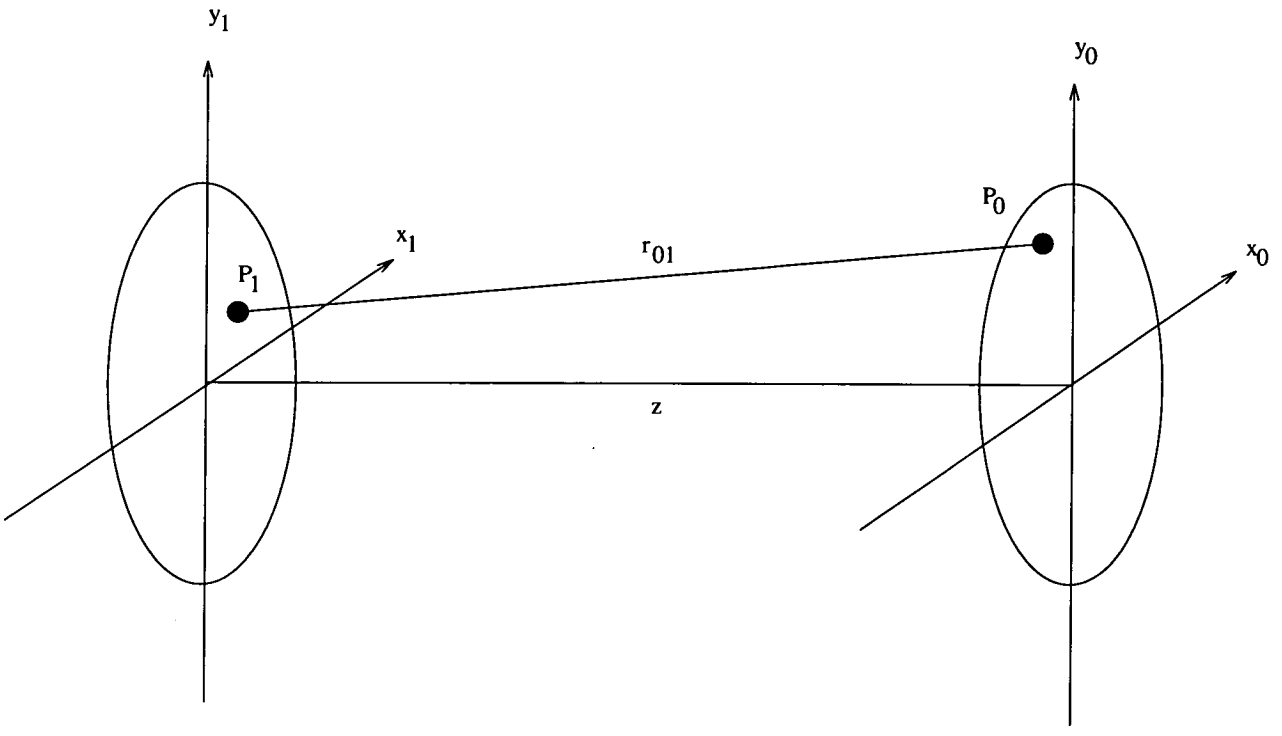


Figure 1.5: The diffraction geometry.

shown in Figure 1.5. Consider two planar screens with rectangular co-ordinate systems (x_1, y_1) and (x_0, y_0) respectively. These two planar screens are parallel to each other with a normal distance denoted by z. P_1 and P_0 are two points located in the plane (x_1, y_1) and plane (x_0, y_0) respectively separated by a distance denoted by r_{01}.

Using the mathematical expression of the Huygens-Fresnel principle, the field amplitude at point (x_0, y_0) can be written as [4,5],

$$U(x_0, y_0) = \int\int h(x_0, y_0; x_1, y_1) U(x_1, y_1) dx_1 dy_1 \tag{1.10}$$

where

$$h(x_0, y_0; x_1, y_1) = \frac{1}{j\lambda} \frac{\exp[jk|\vec{r}_{01}|]}{|\vec{r}_{01}|} \cos(\vec{n}, \vec{r}_{01}) \tag{1.11}$$

$\cos(\vec{n}, \vec{r}_{01})$ represents the cosine of the angle between the outward normal vector $\vec{n}$ and the vector $\vec{r}_{01}$ joining P_0 to P_1, λ is the optical wavelength, $j = \sqrt{-1}$ and k is the wave number which is equal to $\frac{2\pi}{\lambda}$.

The approximation used is based on the assumption that the distance z between two

parallel planes is much greater than the maximum linear dimension of the plane (x_1, y_1) and (x_0, y_0). In this case,

$$\cos(\vec{n}, \vec{r}_{01}) \approx 1 \tag{1.12}$$

and the quantity $|\vec{r}_{01}|$ in the denominator of Eq (1.11) do not differ significantly from z, ie.,

$$\frac{1}{|\vec{r}_{01}|} = \frac{1}{z} \tag{1.13}$$

Hence $h(x_0, y_0; x_1, y_1)$ can be approximated as,

$$h(x_0, y_0; x_1, y_1) \approx \frac{1}{j\lambda z} \exp(jk|\vec{r}_{01}|) \tag{1.14}$$

Note that $|\vec{r}_{01}|$ in the exponent cannot be replaced simply by z, for the resulting errors are multiplied by a very large number k and consequently can generate phase errors much greater than 2 π radians.

Further simplification can be accomplished by adopting certain approximations to the quantity $|\vec{r}_{01}|$ in the exponent. The distance $|\vec{r}_{01}|$ is given exactly by,

$$\begin{aligned} |\vec{r}_{01}| &= \sqrt{z^2 + (x_0 - x_1)^2 + (y_0 - y_1)^2} \\ &= z\sqrt{1 + (\frac{x_0 - x_1}{z})^2 + (\frac{y_0 - y_1}{z})^2} \end{aligned} \tag{1.15}$$

Using the binomial expansion approximation to the square root,

$$\sqrt{1+b} = 1 + \frac{1}{2}b - \frac{1}{8}b^2 + \cdot \tag{1.16}$$

for $|b| < 1$ and retaining only the first two terms of its expansion, $|\vec{r}_{01}|$ can be written as,

$$|\vec{r}_{01}| \approx z[1 + \frac{1}{2}(\frac{x_0 - x_1}{z})^2 + \frac{1}{2}(\frac{y_0 - y_1}{z})^2] \tag{1.17}$$

This assumption, which is referred to as the Fresnel approximation, allows the field at plane (x_0, y_0) to be written as,

$$\begin{aligned} U(x_0, y_0) &= \frac{\exp[jkz]}{j\lambda z} \exp[j\frac{k}{2z}(x_0^2 + y_0^2)] \int\int U(x_1, y_1) \\ &\quad \exp[j\frac{k}{2z}(x_1^2 + y_1^2)] \exp[-j\frac{k}{z}(x_0 x_1 + y_0 y_1)] dx_1 dy_1 \end{aligned} \tag{1.18}$$

Thus, apart from multiplicative amplitude and phase factors that are independent

of (x_1, y_1), the function $U(x_0, y_0)$ may be found from Fourier transforming $U(x_1, y_1)$ $exp[j\frac{k}{2z}(x_1^2 + y_1^2)]$. In other words, it is a parabolic approximation to a wavefront.

The diffraction pattern calculations can further be simplified if more restrictions are adopted. If the approximation that

$$z >> \frac{k(x_1^2 + y_1^2)_{max}}{2} \tag{1.19}$$

is valid, then the quadratic phase factor can be approximated by unity, ie.,

$$\exp[j\frac{k}{2z}(x_1^2 + y_1^2)] \approx 1 \tag{1.20}$$

Then $U(x_0, y_0)$ can be rewritten as,

$$\begin{aligned} U(x_0, y_0) &= \frac{\exp[jkz]\exp[j\frac{k}{2z}(x_0^2 + y_0^2)}{j\lambda z} \\ &\int\int U(x_1, y_1)\exp[-j\frac{k}{z}(x_0x_1 + y_0y_1)]dx_1dy_1 \end{aligned} \tag{1.21}$$

This is the Fraunhofer diffraction approximation. Apart from the multiplicative factors preceding the integral, the function $U(x_0, y_0)$ is simply the Fourier transform of $U(x_1, y_1)$. It is thus a planar approximation to a wavefront. It should be noted that the distance separating the Fresnel and the Fraunhofer regions, and thus the quadratic and the planar approximations to a wavefront, is called the Rayleigh distance [6]. This is defined as,

$$z_R = \frac{D^2}{\lambda} \tag{1.22}$$

where D is the size of the aperture.

1.3 Zernike Polynomial

Because most telescopes and lenses have circular apertures, analysis in polar co-ordinate is simpler and easier than in Cartesian co-ordinate. Zernike polynomials provide a set of orthogonal polynomials defined on a unit circle that are separable in both angle and radius. In this thesis, the definition suggested by Noll [7] is used in which the Zernike

polynomials are defined as,

$$\begin{aligned} Z_{even\,i} &= \sqrt{n+1}\,{R_n}^m(r)\sqrt{2}\cos(m\theta) \quad m \neq 0 \\ Z_{odd\,i} &= \sqrt{n+1}\,{R_n}^m(r)\sqrt{2}\sin(m\theta) \quad m \neq 0 \\ Z_i &= \sqrt{n+1}\,{R_n}^0(r) \quad m = 0 \end{aligned} \tag{1.23}$$

where

$${R_n}^m(r) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s(n-s)!}{s![(n+m)/2-s]![(n-m)/2-s]!}\, r^{n-2s} \tag{1.24}$$

The values of n and m are always integers and satisfy the relationship,

$$n - m = \text{even} \tag{1.25}$$

and $m \leq n$. The index i is a mode-ordering number which is a function of both n and m.

The first few Zernike polynomials together with their common optical descriptions are shown in Table 1.1. One of the advantages of the Zernike polynomials is that they form an orthogonal set of basis functions for a circular telescope aperture. Mathematically, this property is expressed as,

$$\frac{1}{\pi}\int\int_{aperture} Z_i(r,\theta)Z_j(r,\theta)r\,dr\,d\theta = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{1.26}$$

where the integration is performed over the aperture.

Zernike polynomials have been used widely in the field of astronomical imaging, especially in the representation of the wavefront aberration with an assumption of Kolmogorov turbulence as discussed in Chapter 4.

1.4 Optimization Techniques

Optimization problems arise frequently in many different areas. For example, one might want to minimize the material used to make a product or one might want to maximize the profit subject to certain constraints. By focusing attention on a single objective

Radial degree n	Azimuthal frequency m	Zernike polynomials	Descriptions
0	0	$Z_1 = 1$	constant
1	1	$Z_2 = 2rcos\theta$ $Z_3 = 2rsin\theta$	Tilts
2	0	$Z_4 = \sqrt{3}(2r^2 - 1)$	defocus
2	2	$Z_5 = \sqrt{6}r^2 sin2\theta$ $Z_6 = \sqrt{6}r^2 cos2\theta$	astigmatism
3	1	$Z_7 = \sqrt{8}(3r^3 - 2p)sin\theta$ $Z_8 = \sqrt{8}(3r^3 - 2p)cos\theta$	pure coma
3	3	$Z_9 = \sqrt{8}r^3 sin3\theta$ $Z_{10} = \sqrt{8}r^3 cos3\theta$	Zero curvature coma
4	0	$Z_{11} = \sqrt{5}(6r^4 - 6r^2 + 1)$	spherical
4	2	$Z_{12} = \sqrt{10}(4r^4 - 3r^2)cos2\theta$ $Z_{13} = \sqrt{10}(4r^4 - 3r^2)sin2\theta$	5^{th} order astigmatism
4	4	$Z_{14} = \sqrt{10}r^4 cos4\theta$ $Z_{15} = \sqrt{10}r^4 sin4\theta$	

Table 1.1: The first few Zernike polynomials using Noll's definition together with their common optical descriptions.

function designed to model a particular problem and quantify its performance, unknown parameters can be found by optimizing this objective function. This is based on the observations or the measurements available, probably also subject to some constraints limited by the resources.

There are many techniques that have been developed to solve this kind of problems. This section briefly describes three frequently used optimization techniques, they are the maximum likelihood method, the descent method and the projection method.

1.4.1 Maximum Likelihood Method

Maximum likelihood method is based on finding the unknowns that are most likely to have created the observations or the measurements available [8, 9]. For example, consider a discrete random variable X whose probability density function, p(x), is a function of a single parameter Θ. If n independent samples produce the observations $x_1, x_2, \ldots, x_n$, then the probability that this occurs is given by,

$$l = p(x_1)p(x_2)\ldots p(x_n) \tag{1.27}$$

l is commonly known as the likelihood function. Since $p(x_i)$ depends on Θ, the function l also depends on $x_1, x_2, \ldots, x_n$ and Θ. As $x_1, x_2, \ldots, x_n$ are observations or the measurements available, the function l would be a function of Θ only. The basic idea of the maximum likelihood method is to find Θ that is most likely to have created the values $x_1, x_2, \ldots, x_n$. We choose an approximation for the unknown value of Θ for which the likelihood function l is as large as possible.

As the likelihood function contains a total of n terms multiplied together, the mathematics can often be simplified if one takes log on both sides of Eq (1.27), ie.,

$$log(l) = logp(x_1) + logp(x_2) + \ldots + logp(x_n) \tag{1.28}$$

$log(l)$ is called the log likelihood function. Because logarithm is a monotonically increasing function of its argument, that means the particular value of Θ that maximizes the likelihood function will also maximize its log likelihood function.

A particular relevant example to astronomical imaging is the Poisson distribution since

it models light detection. The probability density is given as,

$$Pr(X = x) = \frac{e^{-\mu}\mu^x}{x!} \tag{1.29}$$

where μ is the unknown mean and at the same time the unknown variance of the underlying distribution that has to be found. The joint probability or the likelihood function for n independent observations $x_1, x_2, \ldots, x_n$ can be written as,

$$l = \prod_{i=1}^{n} \frac{e^{-\mu}\mu^{x_i}}{x_i!} \tag{1.30}$$

where $\prod_{i=1}^{n}$ is used to denote the multiplication operation for i from 1 to n. The log likelihood function can be written as,

$$\begin{aligned} L &= \sum_{i=1}^{n} ln(\frac{e^{-\mu}\mu^{x_i}}{x_i!}) \\ &= \sum_{i=1}^{n} [-\mu + x_i ln\mu - ln(x_i!)] \end{aligned} \tag{1.31}$$

Note that since the observations $x_1, x_2, \ldots, x_n$ are fixed, the term, $ln(x_i!)$, is a constant and does not affect the maximization. Thus the log likelihood function can simply be expressed as,

$$L = \sum_{i=1}^{n} [-\mu + x_i ln\mu] + AT \tag{1.32}$$

where AT is another term that does not affect the maximization. Mathematically, the maximization can be done by equating the derivative of L with respect to μ to be zero. This gives

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1.33}$$

in which μ is found to be the expected value of all the observations.

The situation can, however, become very complicated as the number of unknowns increases. Practical situations involving images can have many thousands of unknowns. In this case a direct and analytic solution is not usually possible and an iterative means of maximising the likelihood function must be employed.

The Expectation Maximization (EM) algorithm is a general iterative method to compute the maximum likelihood estimates [10,11]. It has been shown to be applicable to many signal processing applications [12–15]. In the terminology of the EM algorithm,

there are two types of data: complete and incomplete. Complete data (CD) refers to all the points in the original data space. Incomplete data (ID) refers to all the points in a space where observations or measurements are made. Hence ID , denoted by Y, is related to the CD, denoted by X, through a non-invertible many-to-one mapping Γ,

$$Y = \Gamma(X) \tag{1.34}$$

Let θ denotes a vector of parameters that is needed to be estimated. Let $p(Y;\theta)$ and $p(X;\theta)$ denote the probability density functions of the ID and CD data respectively. They are related by,

$$p(Y;\theta) = \int_{X(Y)} p(X;\theta)dX \tag{1.35}$$

Because there are two data sets, there are also two log likelihood functions that are important for estimating the parameters θ from the measured data when the EM algorithm is used. They are the log likelihood functions for ID and CD denoted by $L^{ID}(\theta;\hat{\theta}^{(\mathbf{k})})$ and $L^{CD}(\theta;\hat{\theta}^{(\mathbf{k})})$ respectively.

Starting from an initial admissible estimate $\hat{\theta}^{(0)}$ of the parameters, the EM algorithm produces a sequence of estimates $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots, \hat{\theta}^{(k)}, \ldots$ with non-decreasing CD log likelihood function values. Each iteration comprises two steps, the expectation (E) and the maximization (M) steps. The conditional expectation of the log likelihood function of the CD, given the observed ID and $\hat{\theta}^{(k)}$ can be written as,

$$L^{CD}(\theta;\hat{\theta}^{(\mathbf{k})}) = \mathbf{E}[\log \mathbf{p}(\mathbf{X};\theta)|\mathbf{Y};\hat{\theta}^{(\mathbf{k})}] \tag{1.36}$$

The conditional expectation can be shown to be,

$$L^{CD}(\theta;\hat{\theta}^{(\mathbf{k})}) = \int_{\mathbf{X(Y)}} \log \mathbf{p}(\mathbf{x})\mathbf{p}(\mathbf{x}|\mathbf{Y};\hat{\theta}^{(\mathbf{k})})\mathbf{dx} \tag{1.37}$$

which comprises the E step of the EM algorithm. The M step then determines $\hat{\theta}^{(k+1)}$ parameter by maximizing $L^{CD}(\theta;\hat{\theta}^{(\mathbf{k})})$, ie.,

$$\hat{\theta}^{(k+1)} = \arg_{\max \theta\in\Theta} L^{CD}(\theta;\hat{\theta}^{(\mathbf{k})})\} \tag{1.38}$$

By alternating the E and M steps, the algorithm converges to a stationary point. The

sufficient convergence conditions are,

$$L^{CD}(\theta;\hat{\theta}^{(k)})\text{is continuous in both } \theta \text{ and } \hat{\theta}^{(k)} \tag{1.39}$$

In some cases, it may not be numerically feasible to perform the M step without relaxing the maximization requirement. In this case, $\hat{\theta}^{(k+1)}$ is chosen in such a way that the log likelihood function value increases, rather than being maximized. This variant still ensures that the log likelihood function increases at each iteration cycle [16]. This modification of the maximization requirement leads to the so-called generalized expectation and maximization (GEM) algorithm.

1.4.2 The Descent Method

As stated in the previous section, the classical method to find the turning point of a function is to equate the partial derivatives with respect to all the unknowns to be zero. This method can become very complicated as the number of unknowns increases. Descent method is another common optimization technique that is being used to find the minimum or the maximum point of a function [17–21]. Instead of maximising the likelihood function, the descent method is used to search for the minimum point of a function $f(\mathbf{x})$ (or equivalently the maximum point of a function -$f(\mathbf{x})$). The descent condition, given in Eq (1.40), should be satisfied by the algorithm.

$$f(\mathbf{x_{k+1}}) < \mathbf{f(x_k)} \tag{1.40}$$

for all $k \geq 0$, ie., each iteration should produce a new estimate of $\mathbf{x}$ that is closer to the minimum point of the function than the old estimate. The algorithm can be expressed as follows:

1. *Test for convergence*
   If the conditions for convergence are satisfied, then the algorithm terminates with $\mathbf{x_k}$ as the solution.

2. *Compute a search direction*
   Compute a non-zero vector, $\mathbf{p_k}$, which is the direction of search.

3. *Compute a step length*
   Compute a positive scalar, α_k, which is the step length so that the following inequality holds,
   $$f(\mathbf{x_k} + \alpha_\mathbf{k}\mathbf{p_k}) < \mathbf{f(x_k)} \tag{1.41}$$

4. *Update the estimate of the minimum*
   Set $\mathbf{x_{k+1}} \leftarrow \mathbf{x_k} + \alpha_\mathbf{k}\mathbf{p_k}$, $k \leftarrow k+1$ and go back to Step 1.

The major issue in the descent method is thus reduced to choosing a good direction of search $\mathbf{p_k}$. A simple way to find the search direction is called the steepest descent search. In this method, the direction of search is chosen to be equal to the local downhill gradient at a particular point P_i, ie.,

$$\mathbf{p_k} = - \bigtriangledown \mathbf{f(P_i)} \tag{1.42}$$

The method of steepest descent is conceptually simple and straightforward. However, there are certain numerical difficulties associated with this method. Since the steepest descent method often "zig-zags" toward an optimum solution, the number of iterations required to locate an extremum accurately may thus be very large. Although there is a convergence proof for the mathematical definition of the algorithm of the steepest descent method, it does not guarantee that a practical implementation of the algorithm will converge in an acceptable number of iterations.

In order to improve the rate of convergence, a more sophisticated choice of the direction of search is needed. In the conjugate gradient method, the direction of search is not chosen to be downhill, but rather in a direction that is constructed to be conjugate to the old gradient, and also to all previous directions transversed. Hence the new direction of search would not 'destroy' the minimization achieved by the previous directions of searches. In mathematical terms, the direction of search can be shown to be,

$$\mathbf{p_k} = \begin{cases} - \bigtriangledown f(P_k) & if\, k = 0 \\ - \bigtriangledown f(P_k) + \beta_{k-1}\mathbf{p_{k-1}} & if\, k \neq 0 \end{cases} \tag{1.43}$$

where

$$\beta_k = \frac{[\bigtriangledown f(P_{k+1})]^T \bigtriangledown f(P_{k+1})}{[\bigtriangledown f(P_k)]^T \bigtriangledown f(P_k)} \tag{1.44}$$

and T denotes the transpose of a vector.

1.4.3 Projection

Projection-based algorithms find applications in many different situations [22–26]. It reformulates the optimization problem so that all requirements or constraints that the unknowns have to satisfy are represented by some mathematical sets [27]. For example, if the unknowns are known to be positive, then the corresponding constraint set would be defined as,

$$\{x : x \geq 0\} \tag{1.45}$$

Or if the unknowns are known to be in a prescribed closed interval $[a, b]$, $a \geq 0, b > 0, a < b$, then the constraint set would be defined as,

$$\{x : 0 \leq a \leq x \leq b\} \tag{1.46}$$

After defining all the constraint sets, the algorithm looks for a solution of the unknowns that satisfies all these constraint sets. Mathematically, if there is m a priori known properties of a signal or an image f, f is forced to lie in the intersection of the m given constraint sets $C_1, C_2, \ldots, C_m$, ie.,

$$f \in \bigcap_{i=1}^{m} C_i \tag{1.47}$$

where the symbol $\bigcap$ is used to denote the intersection of all the sets.

Associated with each constraint set C_i, there is a projection operator P_i $(i = 1, 2, \ldots, m)$. In general, it is called

$$g = P_i h \tag{1.48}$$

the projection of h onto the set C_i if $g \in C_i$ and if

$$||g - h|| = min\, ||y - h|| \tag{1.49}$$

over all $y \in C_i$ [28], [23]. The symbol $||x||$ denotes the Euclidean norm of a vector x.

The concept of convexity plays an important role in the application of this algorithm. A set C is said to be convex if, together with x_1 and x_2, it contains $\mu x_1 + (1 - \mu)x_2$ for all μ, where $0 \leq \mu \leq 1$. The projection operator defined above has a unique point if all C_i are convex sets. When one or more of the C_i are non-convex, there may be a

set of points that satisfies the definition of projection [23].

When all the constraints $C_1, C_2, \ldots, C_m$ are of a convex type, the restoration algorithm is given by,

$$f_{n+1} = P_1 P_2 \ldots P_m f_n \tag{1.50}$$

or its relaxed version,

$$f_{n+1} = T_1 T_2 \ldots T_m f_n \tag{1.51}$$

where

$$T_i = 1 + \lambda_i (P_i - 1) \tag{1.52}$$

for $i = 1, 2, \ldots, m$. λ_i is called a relaxation parameter. The algorithm was shown by Youla [27] to converge provided that the following conditions hold, (i) $C_1, C_2, \ldots, C_m$ are closed convex sets with non-empty intersection. (ii) λ_i is between 0 and 2 for all i from 1 to m.

Chapter 2

Practical Preliminaries

The effects of atmospheric turbulence on the image formation of stars at ground level have been known for a long time. In 1730, Isaac Newton wrote in Opticks about the problem of atmospheric turbulence limitations in astronomy. He wrote [29],

"If the theory of making telescopes could at length be fully brought into practice, yet there would be certain bounds beyond which telescopes could not perform. For the air through which we look upon the stars, is in perpetual tremor; as may be seen by the tremulous motion of shadows cast from high towers, and by the twinkling of the fix'd stars. But these stars do not twinkle when viewed through telescopes which have large apertures. For the rays of light which pass through divers parts of the aperture, tremble each of them apart, and by means of their various and sometimes contrary tremors, fall at one and the same time upon different points in the bottom of the eye, and their trembling motions are too quick and confused to be perceived severally. And all these illuminated points constitute one broad lucid point, composed of those many trembling points confusedly and insensibly mixed with one another by very short and swift tremors, and thereby cause the star to appear broader than it is, and without any trembling of the whole. Long telescopes may cause objects to appear brighter and larger than short ones can do, but they cannot be so formed as to take away that confusion of the rays which arises from the tremors of the atmosphere. The only remedy is a most serene and quiet air, such as may perhaps be found on the tops of the highest mountains above the grosser clouds."

Since then, researchers have investigated many methods to improve the quality of the image that can be obtained by ground based instruments. The basic setting in which they have worked is discussed in Section 2.1. It should be noted that even without the presence of atmospheric turbulence, the resolution is limited by diffraction effects. These effects cannot be removed and are a fundamental limit set by the size of the aperture. Section 2.2 describes how turbulence arises and the effects of the turbulence on the quality of the image observed at ground level. In particular the Kolmogorov turbulence model commonly used to describe the turbulence is presented.

Section 2.3 gives a mathematical model concerning the astronomical objects and the images observed at ground level. For all methods that are used to reconstruct the diffraction limited objects from the images detected at ground level, a problem of uniqueness arises. In section 2.4, the uniqueness of reconstructions from blurred one-dimensional signals and two-dimensional images is discussed. In fact it can be seen that the uniqueness arguments for one-dimensional signals and two-dimensional images differ significantly.

2.1 Astronomical Setting

In the field of astronomy, the object of interest is the many radiating objects in the sky. Light from these astronomical objects propagates through the atmosphere and is focussed onto the image plane through the telescope. The telescope is a device which collects as much radiation from astronomical objects and puts it in as sharp(small) an image as possible. A basic astronomical setting is shown in Figure 2.1.

In astronomical imaging application, the propagation distance is very large and the Fraunhofer approximation (Section 1.2) which requires,

$$\begin{aligned} z &> \frac{k(x'^2+y'^2)_{max}}{2} \\ z_o &>> \frac{k(x'^2+y'^2)_{max}}{2} \end{aligned} \tag{2.1}$$

holds in this case. The distance from pupil to the image plane is denoted by z and z_o represents the distance between the object plane and the pupil plane. (x, y) is an arbitrary point in the image plane and (x', y') is an arbitrary point in the pupil plane.

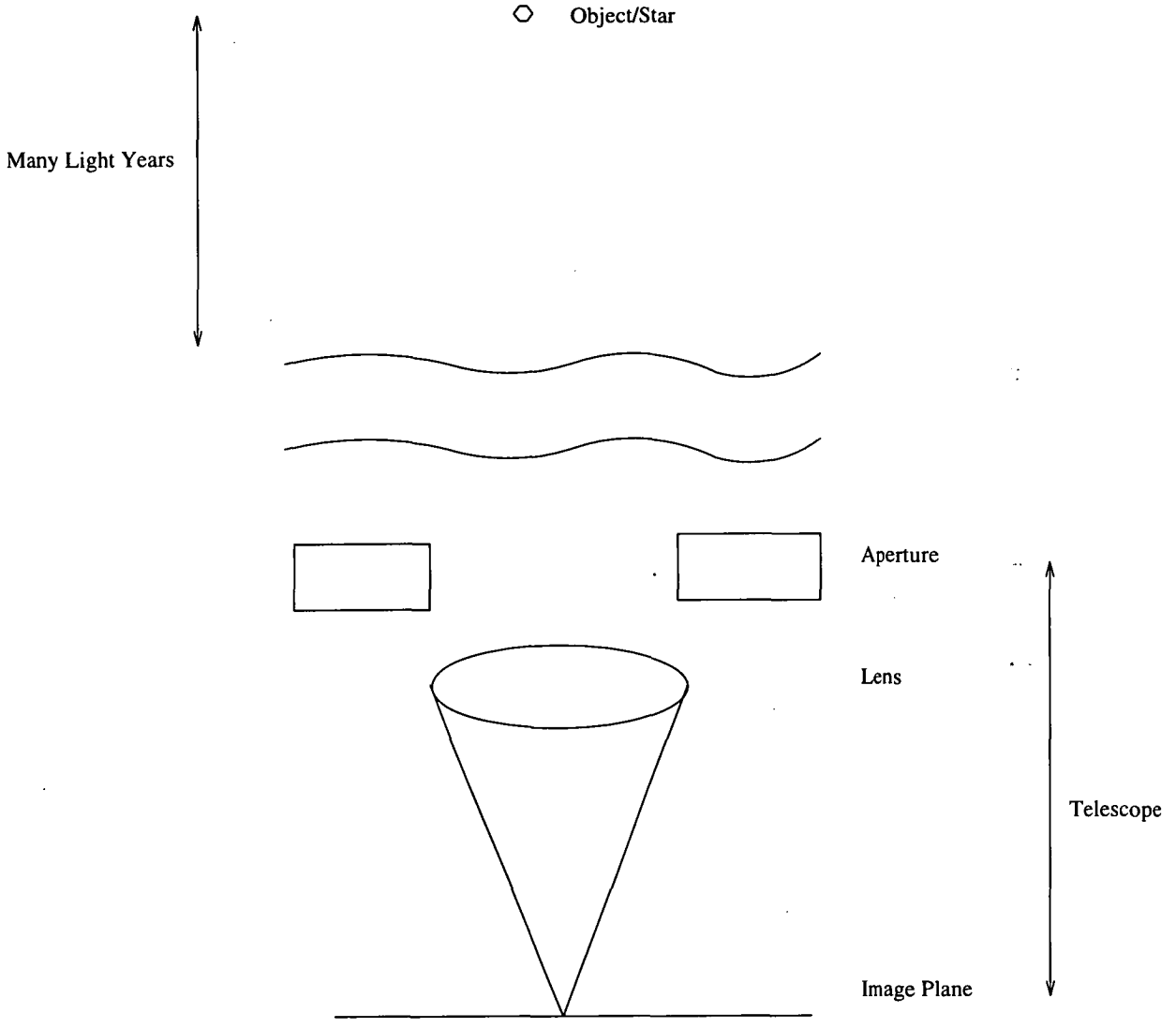


Figure 2.1: A basic astronomical setting.

The field at the image plane for Fraunhofer approximation is basically proportional to the Fourier transform of the field in the pupil plane (Section 1.2), ie.,

$$U(x,y) \propto FT[U(x',y')] \tag{2.2}$$

This is a fundamental equation in the field of astronomical imaging. This equation poses an important physical limitation on the quality of the image that can be observed at ground level. In practice, the telescope aperture is spatially limited because of the finite size of the aperture. As such only a certain portion of the frequency components generated by an object are intercepted by the aperture. The components not intercepted by the aperture are precisely those generated by the high frequency spectral components of the object. As the high frequency components contain the fine details about the object, an aberration-free pupil will not produce the original object, but a smoothed version of the original object. It is generally called a diffraction-limited image. A mathematical model to describe this effect can be written in Fourier domain as,

$$G(u,v) = F(u,v)A(u,v) \tag{2.3}$$

or in image space as,

$$g(x,y) = f(x,y) \odot a(x,y) \tag{2.4}$$

where $A(u,v)$ is a low pass filter that represents the effect of the finite aperture size, $f(x,y)$ is the 'perfect' object and $g(x,y)$ is the observed diffraction-limited image. For a circular aperture with diameter D, the highest spatial frequency that the telescope aperture can intercept is [6],

$$f_{cutoff} = \frac{D}{\lambda} \tag{2.5}$$

where λ is the wavelength. Hence if the diameter of the aperture is large, then many high frequency components can be passed to the imaging system which results in a sharper image.

One way to quantify the sharpness of an image produced by a of telescope is to use resolution. It is defined as the ability of the telescope to separate two closely spaced point sources [4]. The resolution of a perfectly figured, optical telescope is limited by diffraction to be,

$$1.22\frac{\lambda}{D} \quad (in\ radians) \tag{2.6}$$

Thus the resolution improves as the size of the telescope increases.

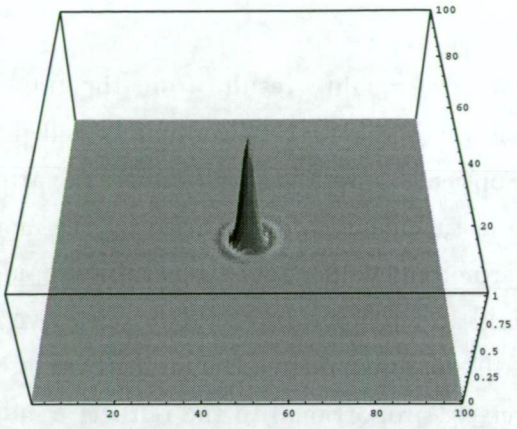


Figure 2.2: The image of a point source object which was observed through a circular telescope aperture under no atmospheric turbulence effect.

A point source observed through a circular aperture forms a diffraction limited image commonly known as Airy pattern. A typical picture of the Airy pattern is shown in Figure 2.2. It consists of a central lobe of this pattern known as Airy disc which has an angular extent approximately equal to $\frac{\lambda}{D}$.

2.2 Problems In Astronomical Images

Theoretically, a large telescope should produce a better quality image than a small telescope because as can be seen in Eq (2.5), more high frequency components are intercepted by the telescope aperture. However, even with a small telescope, the image of a star can be observed to be shifting and scintillating. With a large telescope, the image becomes enlarged and blurred [30]. This problem occurs because the atmosphere also introduces distortion to the imaging system which affects the quality of the image that can be obtained at ground level. The atmosphere limits the resolution of a large telescope to that of a much smaller instrument, thereby degrading the image predicted by theory.

2.2.1 How Turbulence Arises

The degrading effect on image quality results from the transient optical properties of the atmosphere in the line of sight. It is commonly called the seeing problem by the astronomers. These optical properties vary because the atmosphere is a randomly inhomogeneous medium in turbulent motion caused by the movement of wind. Wind leads to the mixing of warmer and colder air by generating rapidly varying 'air bubbles' or 'cells' of different sizes and temperatures. The refractive index of air decreases approximately 1 part in 10^5 for every degree Kelvin increase in temperature [31]. Since the speed of light is inversely proportional to the refractive index, light waves passing through a medium of varying refractive index become distorted, and parts of an incident wavefront travel faster than the other parts of the wavefront. This results in a surface of a constant phase being transformed into a continually deformed wavefront. Although the effect of the atmosphere may be quite small locally, its effects are integrated through the long air path of the starlight before reaching the telescope and thus disturbances can be accumulated considerably.

2.2.2 Statistics Of Turbulence

The atmosphere can be described as a conglomeration of blobs, or seeing cells. These have an average size denoted by Fried's parameter. This is also called the coherence diameter of the atmosphere, and is universally designated as r_0. The phase delay across each blob is roughly constant, but is assumed random with respect to other blobs. More precisely, the Fried's parameter also defines the size of the aperture that can form a diffraction limited image [32]. When the telescope diameter is smaller than r_0, a diffraction limited image can be formed and the resolution of the image is limited entirely by the telescope. But when the telescope diameter is greater than r_0, a blurred image whose size is much larger than the diffraction limit of the aperture is formed and the resolution of the image is limited by the atmosphere [6]. Consequently, a large r_0 value is associated with good seeing conditions whereas small r_0 value means poor seeing conditions. Typical values of r_0 range from a few centimeter up to 20cm or more at the best mountaintop astronomical observatories. r_0 is also a function of wavelength. It increases as $\lambda^{6/5}$, implying that the seeing is better at longer wavelengths.

parameter	symbol	typical values
blob size	r_0	10 cm
time while atmosphere frozen	t_s	10 ms
redistribution time of atmosphere	t_r	500 ms
isoplanatic patch	θ	10 arc second

Table 2.1: Typical values of atmospheric distortion parameters for mid-band visible light which is about 500 nm wavelength. Note that these parameters can vary considerably with location and time.

The Fried's parameter describes the static nature of the atmospheric turbulence. In addition, there are two parameters that are important to describe the time variation of the atmospheric turbulence. The first one is the redistribution time denoted as t_r. It models the time taken for the atmospheric turbulence to change completely. The other one is the correlation time denoted as t_s. It models the time interval when the atmosphere can be considered as effectively stationary. In other words, the atmosphere is effectively 'frozen' for a time t_s, but is completely changed in a time t_r. Since the atmosphere is in time varying, the image formed by a telescope is also time varying, although for a short period of time (less than t_s), the image can be considered effectively frozen. Typical values of atmospheric distortion parameters for mid-band visible light are shown in Table 2.1.

2.2.3 Effects Of Turbulence

The atmospheric turbulence affects both the magnitude and the phase of the wavefront in the telescope aperture. The effect of the aberration is to spread the available light from the astronomical source over a wide area. The most significant effect of the atmo-

sphere on the quality of image formed is the random phase aberration introduced in the optical path of the telescope. This is because phase information is more important than the magnitude information (Section 1.1). Hence many of the different strategies which have been proposed to remove the distortions introduced by the turbulence concentrate on the phase correction, and ignore the magnitude correction.

Mathematically, the wavefront perturbations can be described as follows [33],

$$\varphi(x,y,t) - jA(x,y,t) \tag{2.7}$$

where $\varphi(x,y,t)$ denotes the phase distortion of the wavefront, and $A(x,y,t)$ denotes the magnitude fluctuation across the aperture plane. The former leads to a "dancing" of the star at the image plane while the latter is well known as the twinkling of the stars. These aberrations are three-dimensional, they depend on the positions and also are time-varying.

The atmosphere is closer to the telescope aperture compared to the distance between the object and the atmospheric turbulence. For any point in the pupil plane, rays drawn to all points within the object usually pass through a single cell. The diagram is shown in Figure 2.2.3. The angle subtended at the telescope by an average sized seeing cell in the atmosphere is called the isoplanatic patch. For two wavefronts travelling within the same isoplanatic patch, they would suffer almost the same atmospheric disturbance. For two wavefronts travelling through different seeing cells, the atmospheric disturbance they experience would be different.

The angular resolution of a telescope whose performance is limited by the atmospheric turbulence is [34],

$$1.22\frac{\lambda}{r_0} \tag{2.8}$$

and thus the resolution is degraded by a factor of D/r_0 compared to the resolution of a diffraction limited image. Figure 2.4 and Figure 2.5 show a set of typical pictures of a point source object viewing through different aperture sizes. The object is observed through a circular aperture at ground level and is exposed for a time duration no longer than t_s. We can see that the images are in general enlarged, blurred and shifted because of the presence of the atmospheric turbulence. Also, as the size of the aperture increases, the images develop structure in the form of an array of discrete spots, or speckles. Each speckle actually corresponds to a diffraction-limited replica of

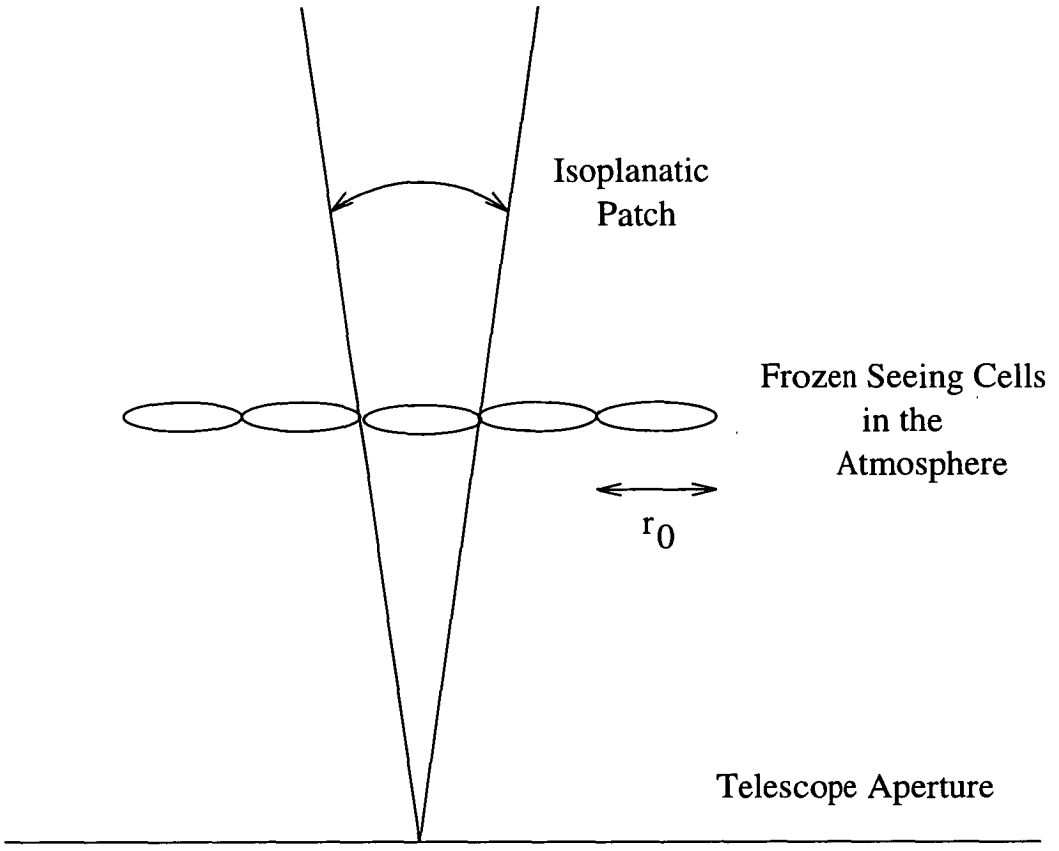
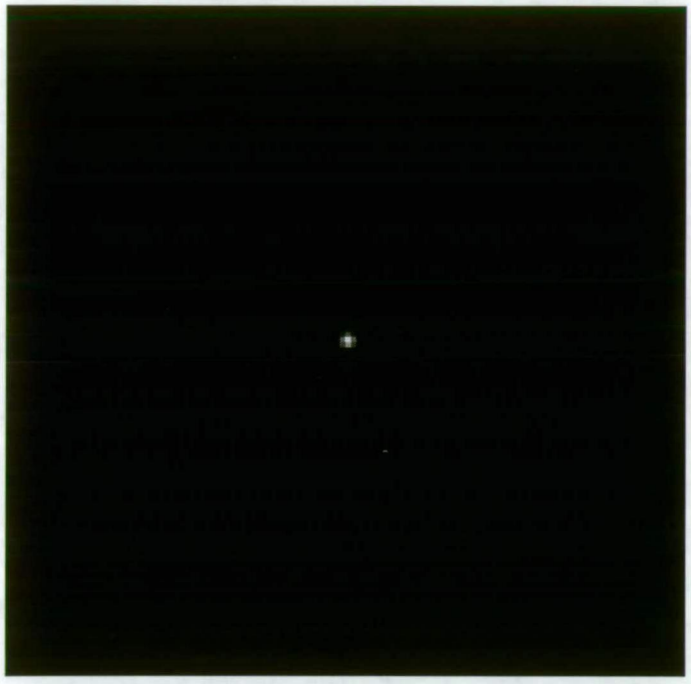
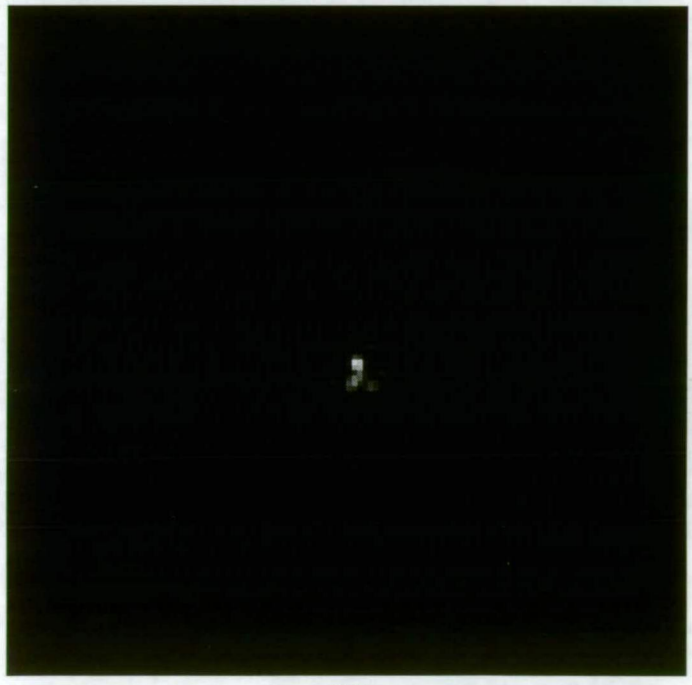


Figure 2.3: The atmospheric turbulence is represented as a series of frozen seeing cells. The angle subtended at the telescope by an average sized seeing cell in the atmosphere is called the isoplanatic patch.

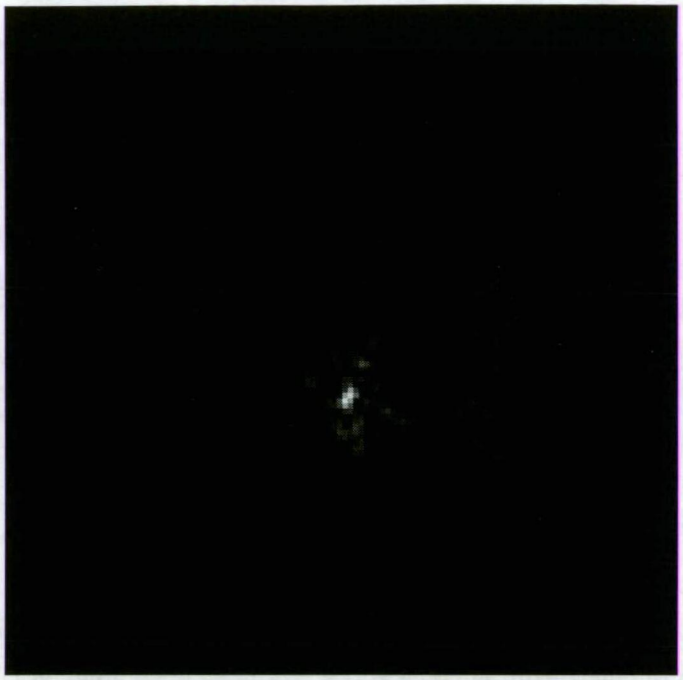


(a)

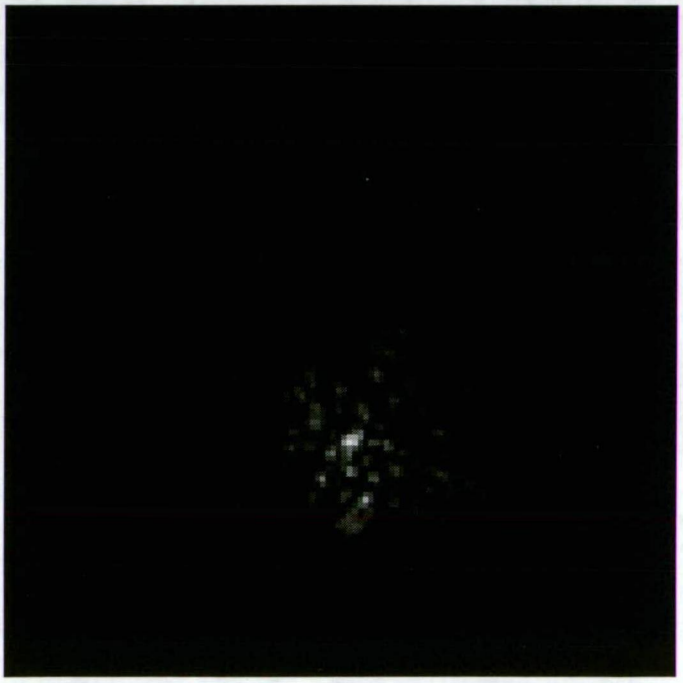


(b)

Figure 2.4: Typical pictures of a point source object observed through the turbulent atmosphere using a circular telescope for different aperture size. The ratio of D to r_0 is set to be (a) 1 and (b) 4.



(a)



(b)

Figure 2.5: Typical pictures of a point source object observed through the turbulent atmosphere using a circular telescope for different aperture size. The ratio of D to r_0 is set to be (a) 8 and (b) 16. The loss of the telescope's resolution is clearly apparent.

the original object [35]. This type of image, commonly known as the short exposure image, preserves information about the original object at spatial frequencies up to the theoretical diffraction limit.

2.2.4 Kolmogorov Turbulence Model

A common model for describing turbulence is the Kolmogorov turbulence model. Ideal Kolmogorov turbulence is both infinite in extent and infinite in detail. It has the unusual property that it 'looks' the same at whatever scale it is viewed. This is in contrast to most conventional mathematical functions which become smooth when viewed in sufficient detail. It is thus fractal in nature [36,37].

One important property of a pure Kolmogorov turbulence is that the wavefront is not stationary, ie., $< \varphi(\vec{r})^2 >$ increases without limit as $|\vec{r}| \rightarrow \infty$ where $\vec{r}$ is a position vector, $\varphi(\vec{r})$ denotes the wavefront surface and $< . >$ is an ensemble average. It is however possible to use a structure function to describe this pure Kolmogorov turbulence model. The structure function is defined as the mean square difference between two points in the wavefront surface separated by a displacement vector $\vec{p}$ [38]. Mathematically, it can be written as,

$$D(p) =< |\varphi(\vec{r}+\vec{p}) - \varphi(\vec{r})|^2 > \tag{2.9}$$

The Kolmogorov turbulence law states that for a wavefront to follow the Kolmogorov turbulence model, the structure function for a plane wave can be written as,

$$D(p) = 6.88(\frac{|\vec{p}|}{r_0})^{5/3} \tag{2.10}$$

The structure function depends solely on $|\vec{p}|$ and r_0.

By examining Eq (2.10), it can also be seen that changing the size of the aperture, ie., either enlarging or reducing the aperture by a factor α, has no effect on the shape of the structure function since,

$$D(\alpha p) = \alpha^{5/3} D(p) \tag{2.11}$$

This implies that, apart form a constant scale factor, the size of the aperture does not affect the statistics of the turbulence.

The other interesting property of the Kolmogorov turbulence model can be seen from its spectrum which can be shown to be [7],

$$\Theta(\vec{k}) = 0.023{r_0}^{-5/3}|\vec{k}|^{-11/3} \tag{2.12}$$

As can be seen from Eq (2.12), the power in the idealized turbulence becomes infinite as $|k| \rightarrow 0$. In spite of this, the structure function shown in Eq (2.10) shows that the variance of the phase fluctuations between any two points of finite aperture are always bounded.

2.3 Mathematical Model

In 1970, Labeyrie [35] proposed to expose an image for a time duration no longer than t_s in order to retain information at spatial frequencies up to the theoretic diffraction limit as discussed in Section 2.2.3. A sequence of N short exposure images could then be recorded. The successive images should be separated in time by at least t_r so that the turbulence can be redistributed and to cause the images to be statistically independent. The n^{th} short exposure image is denoted by $g_n(x, y)$ and can be modelled as,

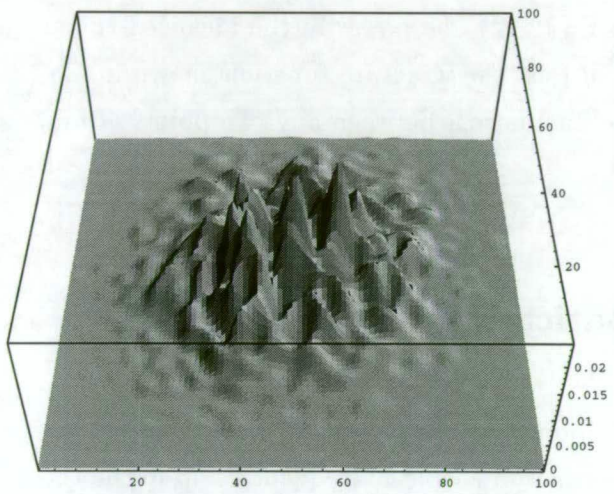
$$g_n(x, y) = f(x, y) \odot h_n(x, y) + c_n(x, y) \tag{2.13}$$

where $f(x, y)$ is the astronomical object, $h_n(x, y)$ is the n^{th} point spread function blurred by both the atmosphere and the telescope, and $c_n(x, y)$ is the noise term. Caluses of this noise include the readout noise of the charged-coupled devices (CCD) and photon noise due to the photon-counting camera used to capture images. Fourier transforming Eq (2.13) gives,

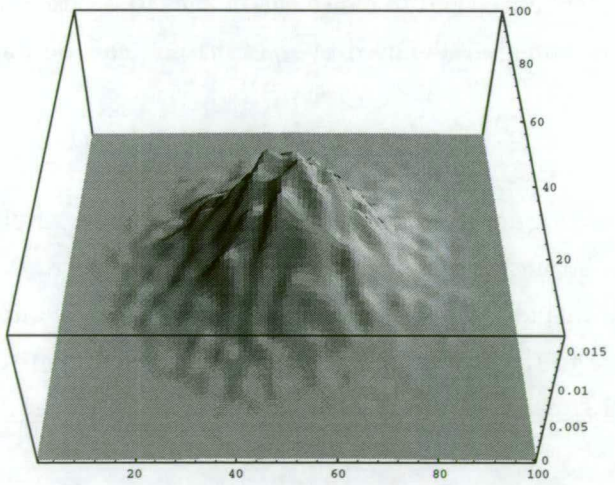
$$G_n(u, v) = F(u, v)H_n(u, v) + C_n(u, v) \tag{2.14}$$

where $H_n(u, v)$ is the optical transfer function.

For most objects of interest, insufficient light ensures that the signal to noise ratio of a single speckle frame is too low to successfully deconvolve on a frame-by-frame basis. The traditional means of increasing the signal to noise ratio, namely averaging the frame, is ineffective in astronomical imaging. The reason for this relates to the phase of



(a)



(b)

Figure 2.6: Typical pictures of a point source object observed through the turbulent atmosphere using different exposure times, (a) a short term exposure image. (b) a long term exposure image produced by averaging 100 short exposure images.

$H_n(u,v)$. Above a certain frequency limit, the phase of $H_n(u,v)$ and hence $G_n(u,v)$ is uniformly distributed between $-\pi$ and π. Hence averaging $G_n(u,v)$ effectively removes these high frequencies components and

$$< G_n(u,v) >= 0 \tag{2.15}$$

at these frequencies. That means if $G_n(u,v)$ is summed over all the frames obtained, the object reconstructed does not have these high frequency components. The reconstructed object thus is smooth compared to the granular appearance of the short exposure image and has a resolution determined by the Fried's parameter. The short exposure image contains more information on small features which is lost in the long exposure image [39]. The typical pictures of long and short exposure images are shown in Figure 2.6. As can be seen, the long exposure image is smoother than the short exposure image and the short exposure image has some high frequency components that are absent in the long exposure image.

2.4 Uniqueness Issues

Following Labeyrie's suggestion, short exposure images that retain infomration at spatial frequencies up to the theoretical limites can be obtained practically. In this section, we consider the uniqueness issue of reconstructing the astronomical objects from these short expsoure images. The mathematical model is given in Eq (2.13). If one can find a method to reconstruct both the object and the point spread function simultaneously from the observed speckle patterns $g_n(x,y)$, then two questions have to be answered. They are,

1. how can one be sure that the estimations are unique or not?
2. under what condition(s) would multiple solutions occur?

It should be noted that $f(x,y)$, $f(x-x_0,y-y_0)$, the image rotated by 180° and $f(-x-x_0,-y-y_0)$ all have the same Fourier magnitude. This fundamental ambiguity does not cause much concern since it is the appearance of the object that is important, not its absolute position. If the above fundamental ambiguities are the only ambiguities, then the object is considered to be unique. Only those solutions having different form are

considered to be ambiguous [40]. We consider two different cases, the one-dimensional signal and the two-dimensional image. In order to simplify the analysis, we consider single frame case only. Thus there is no subscript used below, and g_n and h_n are replaced by g and h respectively.

2.4.1 One-Dimensional Signal

Consider an one-dimensional signal $g(x)$ which is the convolution of $f(x)$ and $h(x)$. The signal, $g(x)$, can be represented by its samples g_m through the Fourier series as,

$$\begin{aligned} g(x) &= f(x) \odot h(x) \\ &= \sum_{m=-M}^{M} g_m \delta(x - x_m) \end{aligned} \tag{2.16}$$

where $x_m = m\triangle x$, and $\triangle x$ is the distance between samples along the x-axis. Fourier transforming Eq (2.16) gives,

$$\begin{aligned} G(u) &= F(u)H(u) \\ &= \sum_{m=-M}^{M} g_m \exp[-j2\pi m u \triangle x] \end{aligned} \tag{2.17}$$

By putting $z = \exp[-j2\pi u \triangle x]$, Eq (2.17) becomes,

$$\begin{aligned} G(z) &= F(z)H(z) \\ &= \sum_{m=-M}^{M} g_m z^m \end{aligned} \tag{2.18}$$

$G(z)$ is generally called the z-transform of $g(x)$ in which $G(z)$ is written as a polynomial. Hence, attempting to deconvolve a one-dimensional signal is equivalent to finding the zeros of the one-dimensional signal, $G(z)$, and then sorting out which zeros correspond to $f(x)$ and which correspond to $h(x)$.

A traditional way to find the zeros is to factorize the one-dimensional signal. It can be advantaged by the presence of the Fundamental Theorem of Algebra. The Fundamental Theorem of Algebra states that any polynomial in one variable of degree 2 or more is reducible over the field of complex numbers and may be expressed as a product of first

order factors. As a result, a function of degree M can always be factored into a product of M first order factors. If deconvolving this M^{th} order function, there are a total of 2^{M-1} different combinations of the zeros if all of them are distinct [41]. As M increases, the total number of combinations can be huge. Hence for a one-dimensional case, the degree of ambiguity is always high and uniqueness is difficult to assure in most cases.

2.4.2 Two-Dimensional Image

Polynomials in more than one variable frequently arise in practice. A two-dimensional image, $g(x,y)$, can always be represented by its samples $g_{m,n}$ through a two-dimensional Fourier series as,

$$\begin{aligned} g(x,y) &= f(x,y) \odot h(x,y) \\ &= \sum_{m=-M}^{M} \sum_{n=-N}^{N} g_{m,n}\delta(x-x_m, y-y_n) \end{aligned} \tag{2.19}$$

where $x_m = m\triangle x$, $y_n = n\triangle y$ and $\triangle x$ and $\triangle y$ represent the spacings between samples in the respective x and y directions. Fourier transforming it gives,

$$\begin{aligned} G(u,v) &= F(u,v)H(u,v) \\ &= \sum_{m=-M}^{M} \sum_{n=-N}^{N} g_{m,n} \exp[-j2\pi(mu\triangle x + nv\triangle y)] \end{aligned} \tag{2.20}$$

By putting $z = \exp[-j2\pi u\triangle x]$ and $w = \exp[-j2\pi v\triangle v]$, Eq (2.20) becomes,

$$\begin{aligned} G(z,w) &= F(z,w)H(z,w) \\ &= \sum_{m=-M}^{M} \sum_{n=-N}^{N} g_{m,n} z^m w^n \end{aligned} \tag{2.21}$$

Again, attempting to deconvolve $g(x,y)$ is equivalent to factorizing $G(z,w)$ and then sorting the zeros to $F(z,w)$ and $H(z,w)$. The result of ambiguity in a one-dimensional case, however, cannot be extended straightforwardly to the functions with two or more variables because of the absence of a fundamental theorem of algebra which enables one-dimensional polynomials to always be factorised [41].

Hayes and McClellan [42] shows that the set of factorable multi-dimensional polyno-

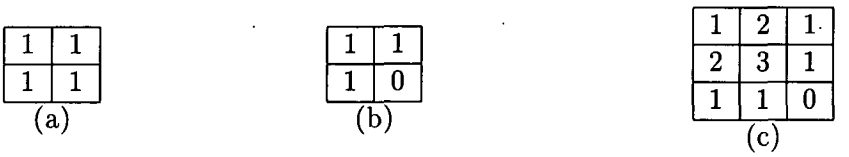


Figure 2.7: An example of a simple convolution between (a) $f(x,y)$ and (b) $h(x,y)$. The resultant convolved image, (c) $g(x,y)$

Figure 2.8: The two convolutional components of $f(x,y)$, (a) $o1$ (b) $o2$

mials is small in the sense that almost all polynomials in two or more variables are irreducible. Hence to most of the practical problems, the deconvolution in a two dimensional case is unique. However, Huiser and Van Toorn [43] have found a condition for a two dimensional deconvolution problem to be non-unique. If the two dimensional object is a convolution of some other images, then there are multiple solutions to the deconvolution problem. The following simple example demonstrating the non-uniqueness situation in the two-dimensional case.

The object, point spread function and their convolution are shown in the Figure 2.7. A difficulty arises in this case because the object itself is a convolution of two parts shown in Figure 2.8 and are denoted as $o1$ and $o2$. It is difficult to assure uniqueness in this case. Without the presence of noise, the image can be expressed as the convolution of $o1(x,y)$, $o2(x,y)$, and $h(x,y)$, ie.,

$$\begin{aligned} g(x,y) &= f(x,y) \odot h(x,y) \\ &= o1(x,y) \odot o2(x,y) \odot h(x,y) \end{aligned} \tag{2.22}$$

Therefore, there are four possible classes of solution that would produce the observed convolution. These are,

1	$\hat{f}(x,y) = k \times o1(x,y) \odot o2(x,y)$	$\hat{h}(x,y) = \frac{h(x,y)}{k}$
2	$\hat{f}(x,y) = k \times \delta(x,y)$	$\hat{h}(x,y) = \frac{1}{k} o1(x,y) \odot o2(x,y) \odot h(x,y)$
3	$\hat{f}(x,y) = k \times o1(x,y)$	$\hat{h}(x,y) = \frac{1}{k} o2(x,y) \odot h(x,y)$
4	$\hat{f}(x,y) = k \times o2(x,y)$	$\hat{h}(x,y) = \frac{1}{k} o1(x,y) \odot h(x,y)$

From a mathematical viewpoint, all of these are equally valid solutions. Thus in this

case, there are four possible pairs of solutions to the blind deconvolution problem which demonstrate the non-uniqueness in a two-dimensional case.

In practical situation, it is possible to reduce the inherent ambiguities nature of the problem. For example, one can obtain an ensemble of differently blurred versions of the object. Because of the time-varying nature of the atmosphere, each version of the image would have a different blurring function but the object remains common for each version. Hence an ensemble of blurred images could effectively eliminate cases (3) and cases (4). Cases (2) is particularly troublesome as it always provides a perfect solution even in the presence of noise. In practice, however, we are not interested in this trivial solution and a number of methods have been proposed to avoid this solution. These methods are discussed in Chapter 6.

Chapter 3

Solutions To The Seeing Problem

There are two general classes of techniques to deal with the distortion associated with the atmospheric turbulence. They are the post-detection processing and the real time processing techniques. The basic principle of the post-detection processing technique is to freeze the atmospheric turbulence by obtaining a sequence of images by using a short term exposure time. This method is founded upon the insight of Labeyrie's idea [35] that high spatial frequency components are preserved in these short exposure images (Section 2.3). An ensemble of these short exposure images is then post-processed in a computer to recover the diffraction-limited object. The second class of technique is the adaptive optics approach, which works in real time. It uses hardware designed to physically compensate for the different atmospheric optical path lengths across the aperture in real time [33]. It, as yet does not compensate for any amplitude distortion of the incoming wavefront.

The advantage of a real time compensation system over the alternative approach of post-detection processing is that an image with improved resolution and a good signal to noise ratio can be obtained in real time. In addtion, further improvement is still possible by subsequent post-detection processing techniques. But the cost associated with implementing the adaptive optics technique can be large.

In Section 3.1, the basic idea of the adaptive optics real time processing technique is presented. In particular, different essential components of the adaptive optics are outlined and the limitations of the performance of the adaptive optics are discussed. Section 3.2 discusses the idea of the speckle interferometry proposed by Labeyrie in 1970 [35]. As only the autocorrelation data is provided by the speckle interferometry, phase information needs to be retrieved. Some of the methods for phase retrieval are outlined in this section.

A technique which combines some elements of adaptive optics with computer processing, called wavefront sensing deconvolution, is discussed in Section 3.3. This technique relies on simultaneously recording of both the distortion information and the image of the astronomical object. The distortion is then removed afterwards by computer.

The techniques discussed above all require the measurements of the distortion information. If this information is unavailable and has to be estimated together with the object, then the problem of the object reconstruction poses a blind deconvolution problem. The most common way to solve the blind deconvolution problem is to use the iterative blind deconvolution loop. It is a modification of the phase retrieval algorithm and it uses some a priori information of both the object and the point spread function to recover the original object [44, 45]. Another technique for solving the blind deconvolution problem is the zero sheet based blind deconvolution technique [46]. This technique is based on the observation that a two-dimensional convolved image has a distinct set of zeros for the object and the point spread function in the absence of noise. These two blind deconvolution techniques are discussed in Section 3.4.

3.1 Real Time Adaptive Optics Approach

The concept of using adaptive optics to solve the seeing problem was originated by H.W. Babcock in 1953 [47]. He proposed a means of compensating the atmospheric turbulence by continually measuring the deviations of rays for all parts of the mirror. This error signal was then amplified and fed back either to directly modify the mirror or apply alternatively as a correction at the conjugate image plane of the mirror. In a latter case, a relatively small ray controlling element consisting of a thin layer of oil covering a reflecting mirror was inserted to control the optical phase shifts in the

reflected light. By passing the part of the wavefronts that had phases in advance of other parts through a thicker film of liquid, the phase distortion caused by the turbulence can be compensated. Although this approach is no longer used, following Babcock's suggestion, it was the forerunner of adaptive optics currently used in both the astronomical and military applications. Indeed it has been shown that at high light levels closed loop corrections of atmospheric disturbances can produce near diffraction-limited images [48].

3.1.1 Basic Idea Of Adaptive Optics

Current adaptive optic systems remove the atmospheric disturbance by inserting an optical component in the passage of light beams. The purpose of inserting the component is to produce a distortion which both spatially and temporally conjugate to the wavefront distortion induced by the atmospheric turbulence. In the case where the wavefront is measured with the sufficient accuracy both spatially and temporally, and in which the control of the optical component is perfect, the phase component of the atmospheric induced distortion can be removed and the telescope will give a diffraction-limited image [29].

The principle of compensation by a counter wavefront can best be understood by considering a simple example shown in Figure 3.1 [29,49]. The wavefront of a beam entering from the left, denoted by (a), is distorted by an aberrator, such as a piece of glass, because of the change of refractive index. As the speed of travel is inversely proportional to the refractive index, some parts of the wavefront travel more slowly than the other parts since the wavefront is retarded by passage through the glass. The resultant wavefront is denoted by (b). After reflecting from the mirror, which is denoted by (c), the wavefront has the same shape but is propagating in the opposite direction. As it transverses the aberrator again, it receives the same retardation as before. The resultant wavefront, denoted by (e), is greatly distorted as it passed through the aberrator twice.

If a plane wavefront is needed after a beam passed through the aberrator twice, there may be a way to alter the surface of the mirror in such a way as to invert the wavefront so that the second passage leaves no residual distortion. The procedure is shown in Figure 3.2. When a bump in the mirror is at the right place and has the right amplitude,

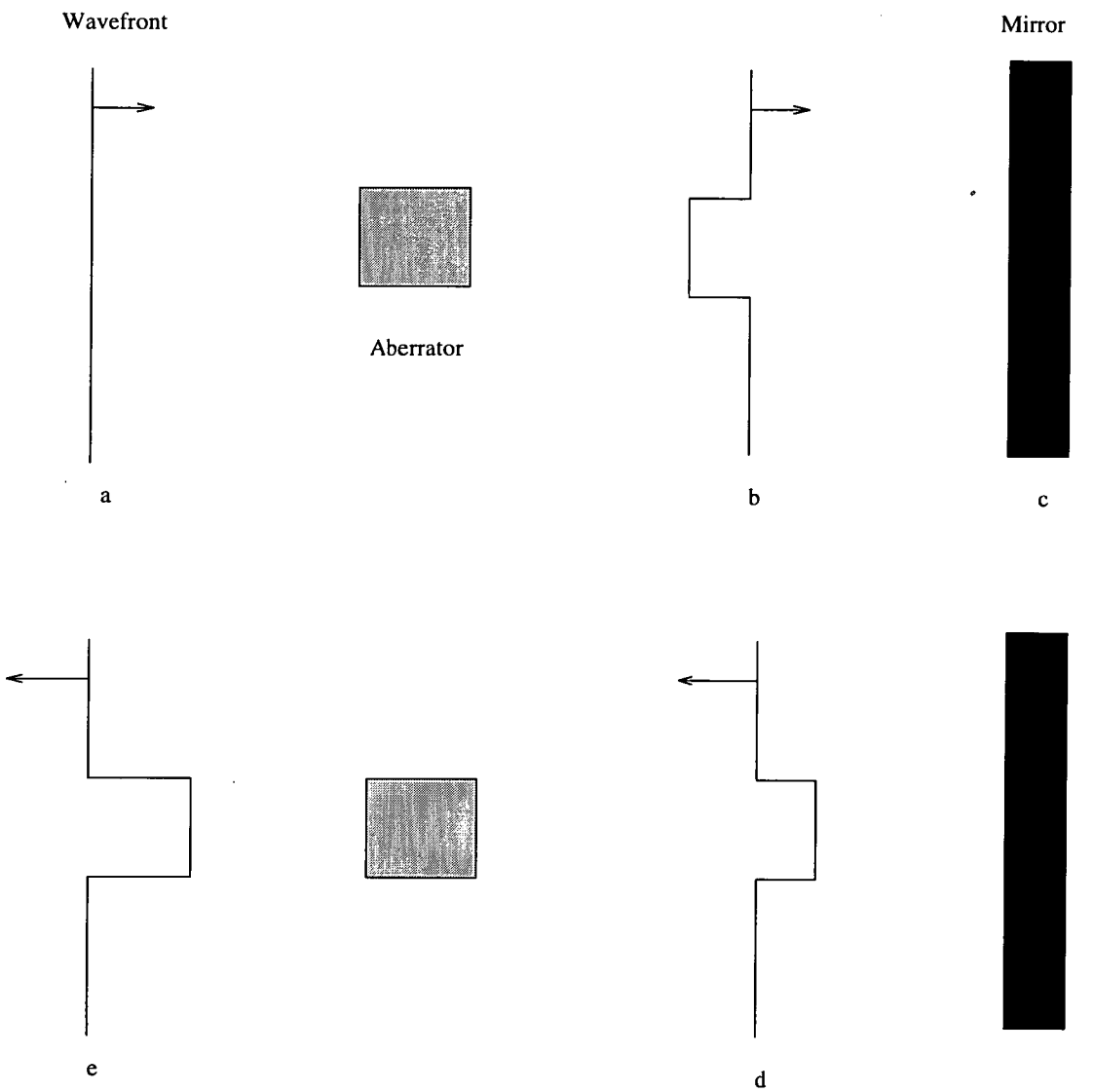


Figure 3.1: A simple example to demonstrate the principle of optical reciprocity. This picture shows the resultant wavefront after travelling through an aberrator.

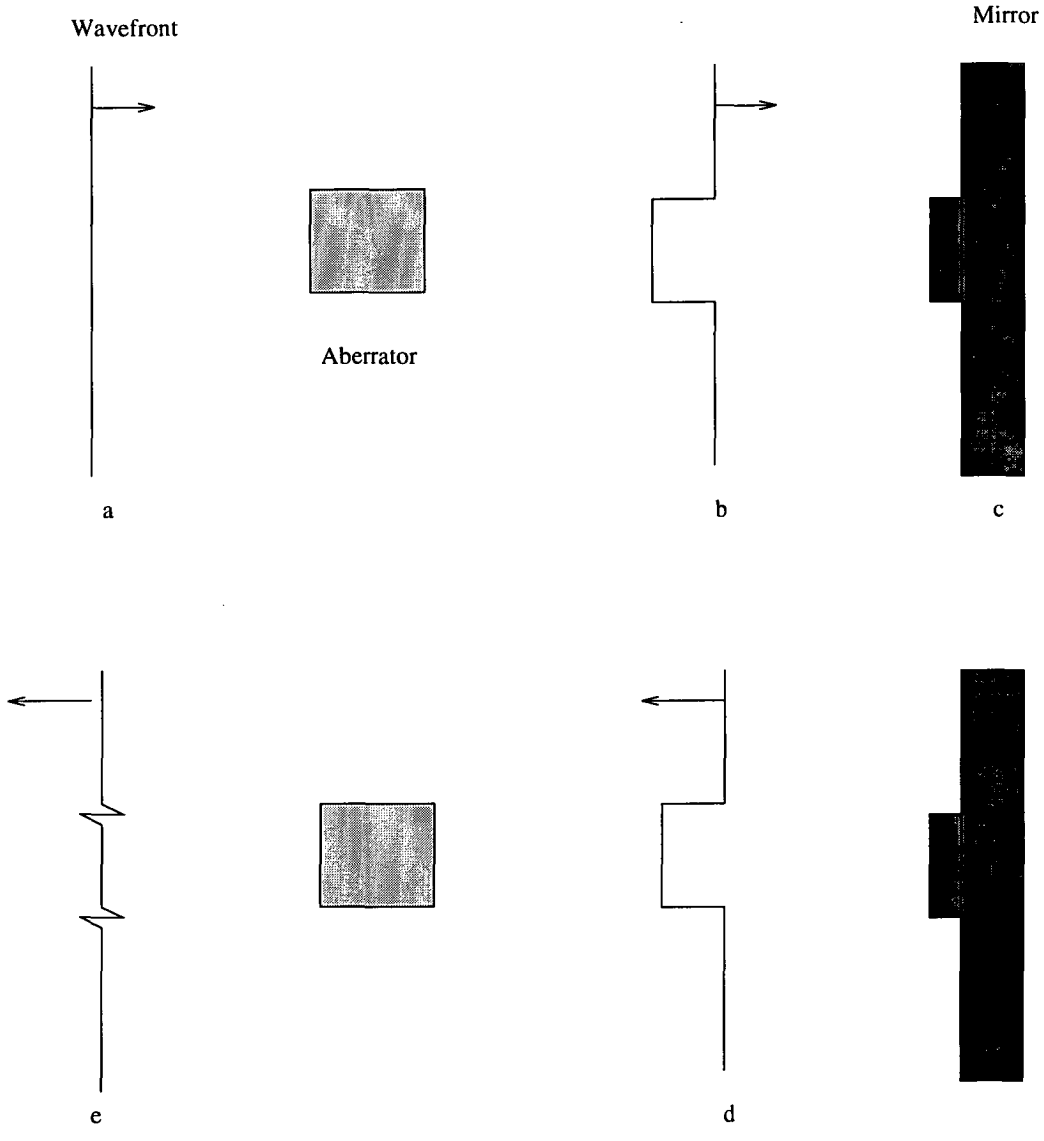


Figure 3.2: The resultant wavefront after travelling through an aberrator but with a correcting mirror.

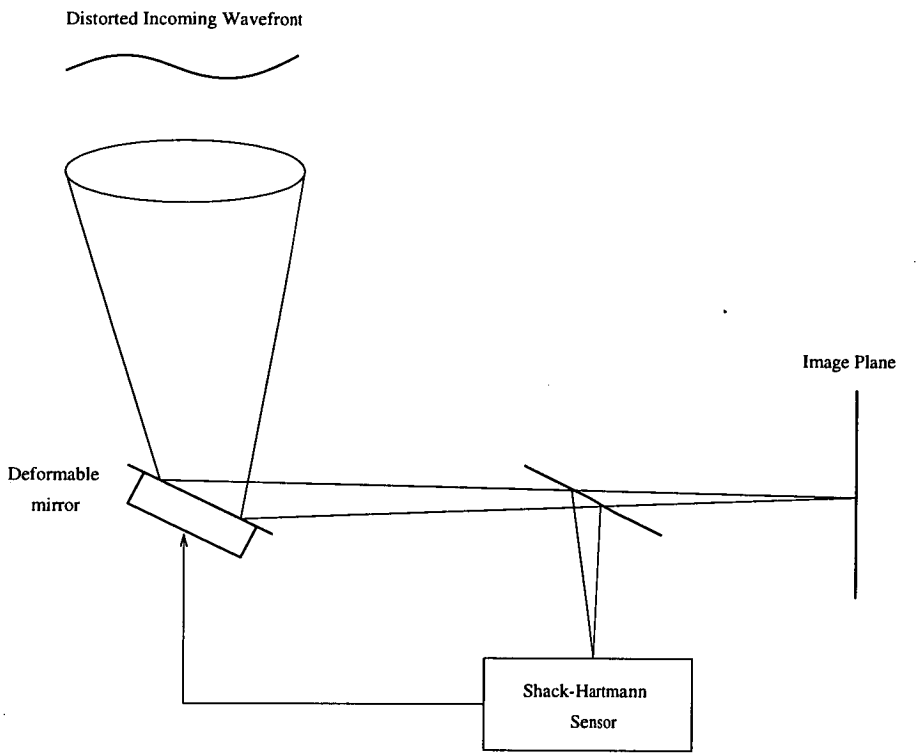


Figure 3.3: A simple adaptive optics system

the leading edge of the wavefront can be reversed. The wavefront, denoted by (d), then passes through the aberrator again, and the final wavefront, (e) is again a plane wave.

Hence in the astronomical imaging, if the distortion can be sensed and the correction is applied at the 'right' place and at the 'right' amount, the distortion introduced by the atmospheric turbulence could be removed, thus making the recovery of the diffraction-limited astronomical object possible.

3.1.2 The Conceptual Components Of An Adaptive Optics System

Figure 3.3 shows a basic closed loop adaptive optics system. It consists conceptually of the following components [50]:

1. a light collecting aperture;
2. a correcting element such as a deformable mirror in which each small part of the surface of the mirror can be moved independently to provide a variation in the

optical path length of the light reflected from that part;

3. a set of electro-mechanical relays to control the movement of the small part of the surface of the mirror;

4. a sensing mechanism which gives information on the spatial distribution of the atmospherically induced phase error, eg, an array of photon detectors that records the places and times of arrival of photons that passed through a lenslet array.

5. a computer that accepts inputs from the sensor and controls the movement of the mirror surface in response to the optical performance of the system at previous times;

A key part of the overall system is an algorithm to estimate the wavefront distortion and convert this information to signals that are fed to the deformable mirror to compensate the distortion.

3.1.3 The Assessment Of The Wavefront Distortion

The sensing of the wavefront distortion information is the first essential stage in compensating the disturbances associated with the atmospheric turbulence in real time. It is, however, difficult to measure the phase of the incoming wavefront at points across the telescope aperture directly. An alternative way to assess the wavefront distortion is to measure the wavefront slope. Consider the configuration in Fig 3.4. For a tilted wavefront, the spot formed in the image plane is displaced in an amount proportional to the tilt of the wavefront. Hence in this way, the mean slope of the wavefront can be calculated from the measurements taken in the image plane.

In order to have a more accurate estimate of the wavefront, the wavefront can be sampled by an array of lenslets. Each lenslet defines a sub-aperture and cuts out a part of the wavefront. Each sub-aperture provides a low-resolution sub-image of the object. Because of the presence of atmospheric turbulence, the distorted wavefront would result in positional variations of the sub-image on the detector. This is the principle of the Shack Hartmann sensor. A diagram of the measurements that can be expected from the sensor is shown in Figure 3.5. When the spots are formed without the effect of the

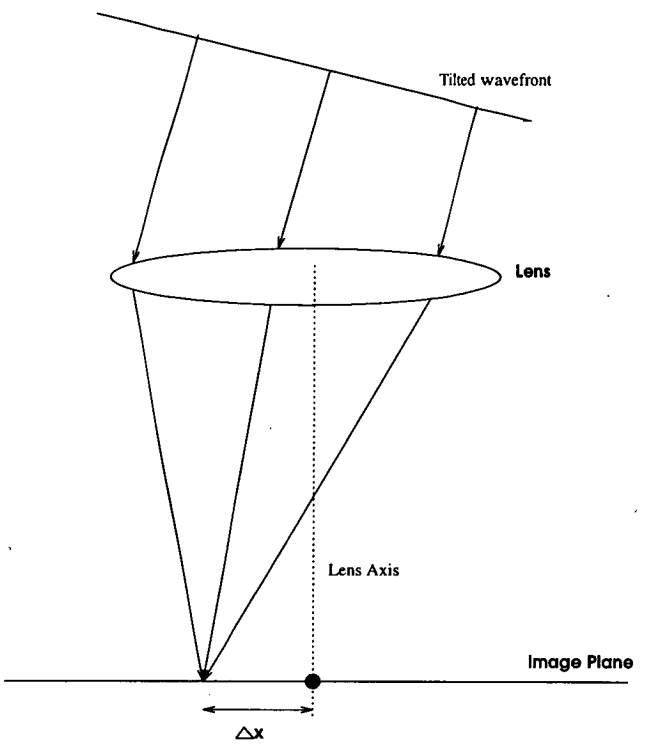
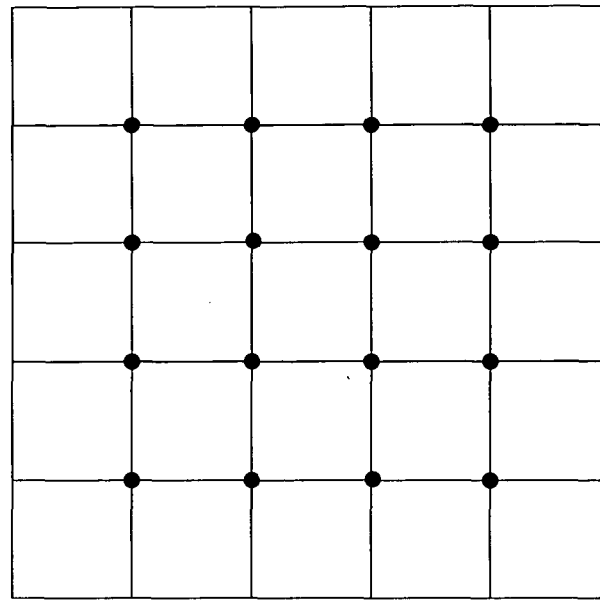
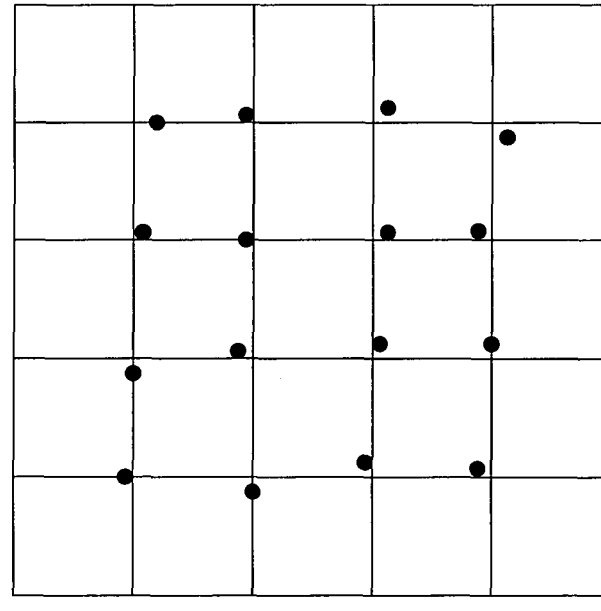


Figure 3.4: An alternative way to measure the wavefront slope is to measure the shift of the spot in the image plane. The tilt of the wavefront is proportional to the displacement of the spot formed.



(a)



(b)

Figure 3.5: Some typical measurements from the Shack-Hartmann sensor, (a) without the effect of turbulence, (b) with the presence of turbulence.

turbulence, they are located in the centroid of each sub-aperture. But when the image is distorted by turbulence, each spot is shifted inside the sub-aperture.

Mathematically, each sub-aperture gives two measurements, x_c and y_c. They are the displacements of the sub-image in two orthogonal directions denoted by x and y. It can be shown that [51],

$$x_c = \frac{\lambda f}{2\pi A_{sa}} \int\int_{sa} \frac{\partial \phi}{\partial u} du dv \tag{3.1}$$

and

$$y_c = \frac{\lambda f}{2\pi A_{sa}} \int\int_{sa} \frac{\partial \phi}{\partial v} du dv \tag{3.2}$$

where A_{sa} is the area of the sub-aperture, f is the focal length of a lenslet and $\frac{\partial \phi}{\partial u}$ is the u-derivative of the plane of the wavefront. By rearranging Eq (3.1) and Eq (3.2), the mean slope in x and y directions can be stated as,

$$\frac{1}{A_{sa}} \int\int_{sa} \frac{\partial \phi}{\partial u} du dv = \frac{2\pi x_c}{\lambda f} \tag{3.3}$$

and

$$\frac{1}{A_{sa}} \int\int_{sa} \frac{\partial \phi}{\partial v} du dv = \frac{2\pi y_c}{\lambda f} \tag{3.4}$$

respectively. Hence the mean slope of the wavefront over a subaperture can be calculated from the displacements. From this information, an estimation of the wavefront can be carried out.

3.1.4 The Estimation Of The Wavefront Distortion

After obtaining a set of wavefront slope measurements, the problem of estimating wavefront phases from these measurements has to be solved. A number of authors [52–56] have addressed this problem. These can be divided into two main classes, namely the zonal and the modal approaches. In the zonal approach, the estimates are the phase values located at the centre of each grid position. In the modal approach, the phase is characterized by an expansion over a set of basis functions. The slope measurement sampling geometry, including the grid positions for the x- and y-slope measurements and the reconstructed phase, is shown in Figure 3.6. The problem is to relate the slope measurement data from the Shack-Hartmann sensor to the phases of the wavefront at the grid points.

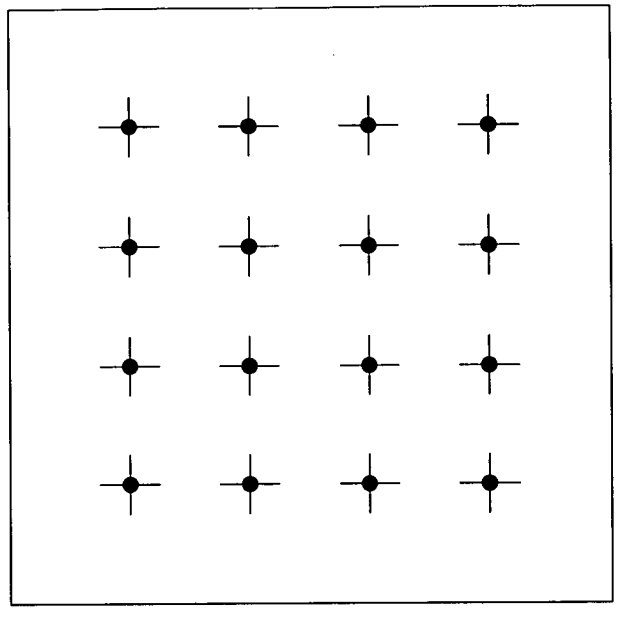


Figure 3.6: A diagram showing the slope measurement sampling geometry. The horizontal dashes indicate positions of x-slope sampling. The vertical dashes are the y-slope sampling positions. The dots are the estimated phase points.

In the zonal estimate, the recursive relationship of the slope measurements and the phases are given by [56],

$$\frac{{S_{i+1,j}}^x + {S_{i,j}}^x}{2} = \frac{\phi_{i+1,j} - \phi_{i,j}}{h} \tag{3.5}$$

$$\frac{{S_{i+1,j}}^x + {S_{i,j}}^y}{2} = \frac{\phi_{i,j+1} - \phi_{i,j}}{h} \tag{3.6}$$

where S^x and S^y are the slope measurements in the x-direction and y-direction respectively, $\phi_{i,j}$ is the phase value at the grid (i,j), N is the number of phase sample points in either x- or y-directions, D is the width of the aperture and h is the distance between two phase measurements, ie., $h = \frac{D}{N}$.

In the modal estimate, the incident wavefront onto the telescope aperture is written in terms of a set of basis functions as,

$$\phi(x,y) = \sum_{i=0}^{\infty} a_i \psi_i(x,y) \tag{3.7}$$

where $\phi(x,y)$ represents the unknown wavefront incident onto the telescope aperture, $\psi_i(x,y)$ is the i^{th} basis function, and a_i is the weighting attached to the i^{th} basis

function. The measurements from the Shack-Hartmann sensor can be expressed as,

$$\frac{\partial \phi}{\partial x_j} = \sum_i a_i \frac{\partial \psi_i(x,y)}{\partial x_j} \tag{3.8}$$

$$\frac{\partial \phi}{\partial y_j} = \sum_i a_i \frac{\partial \psi_i(x,y)}{\partial y_j} \tag{3.9}$$

It can be seen that, no matter which approach one uses, the estimation problem can still be rewritten in matrix notation as,

$$\mathbf{h} = \Theta \mathbf{a} \tag{3.10}$$

where $\mathbf{h}$ is a vector containing all the slope measurements from the Shack Hartmann sensor, Θ is the system matrix and $\mathbf{a}$ is a vector containing the unknown phase values in the zonal approach, and is a vector containing the expansion coefficients in the modal approach. A detailed discussion of the solution of Eq (3.10) is given in Chapter 4.

3.1.5 The Correction Of The Wavefront Distortion

Once the wavefront distortion is known, it is the task of the correcting elements such as deformable mirrors to correct the distortion. The mirrors have to be designed so that their shape can be adjusted to match the instantaneous wavefront distortion well. These mirrors are characterized by [57, 58],

1. their size,
2. the number of actuators which in turn determines the highest spatial frequency of the aberration that the mirror can compensate,
3. the wavefront influence function of each actuator which is the shape of the mirror surface and which defines the spatial spectrum of the correction,
4. the speed at which they can be adjusted,
5. their stroke which is the peak to peak displacement that an actuator can achieve.

There are two common types of deformable mirrors, they are the segmented mirrors and the continuous faceplate mirrors. The segmented mirror consists of many flat plates.

Since each segment is physically separated from the others, segmented mirrors have large degree of freedom of movement and have large capacity to compensate for severe turbulence. However, because each segment is discontinuous to the others, segmented mirrors tend to diffract some of the lights which would affect an image's clarity.

Continuous faceplate mirrors can, in general, work with a reduced number of actuators because they automatically maintain continuity. However, each faceplate is not physically separated which reduces the degree of the freedom of movement. One of the common mathematical models for the shape of the surface of the deformable mirrors is the Gaussian influence function. The actuator response can then be written as,

$$r_j(x,y) \propto exp[-\frac{(x-x_j)^2+(y-y_j)^2}{L_a^2}] \tag{3.11}$$

where x and y specify a point in the plane of the mirror, x_j and y_j specify the actuator position, L_a is the influence radius and r_j is the j^{th} actuator's influence function.

3.1.6 Limitations Of The Performance

In practice, there are many practical limitations that prevent the adaptive optics system from achieving diffraction-limited resolution [58, 59]. Firstly, the wavefront sensor provides only a finite number of measurements, and these measurements are subjected to the noise contamination of the recording process. Also, it requires sufficient photons to make accurate slope measurements. This is particularly critical in astronomy where the most scientific interest is usually associated with the faintest objects.

Secondly, there must be a time delay between the wavefront measurement and its correction. This error becomes significant when the atmospheric turbulence changes very quickly. Also, due to the finite number of degree of freedom, the reconstructed wavefront from the deformable mirror may not exactly match to the actual wavefront.

Finally, as discussed in Chapter 2, most of the corrections are applied to compensate the phase distortion only, and the amplitude distortion is neglected. The residual error in the amplitude distortion limits the the adaptive optics system to a resolution below the diffraction limited resolution.

3.2 Speckle Interferometry

Speckle interferometry was proposed by A. Labeyrie in 1970 [35] as a means of achieving diffraction limited resolution for large telescopes despite the presence of atmospheric turbulence. Labeyrie proposed to use short exposure images as they retain information at spatial frequencies up to the theoretic diffraction limit, a point noted in Section 2.3. Mathematically, one estimates the autocorrelation function or its Fourier transform pair, power spectrum, of the image observed at ground level, $g_n(x, y)$, from an ensemble of short exposure images.

$$< |G_n(u,v)|^2 >= |F(u,v)|^2 < |H_n(u,v)|^2 > \tag{3.12}$$

where $< |H_n(u,v)|^2 >$ is called the speckle interferometry transfer function.

The key feature of speckle interferometry is that the speckle interferometry transfer function, $< |H_n(u,v)|^2 >$, is positive and non-zero for all frequencies up to the diffraction limit of the telescope. This transfer function can be obtained experimentally by observing a single unresolvable star for which all observing conditions are assumed to be identical to those for the object of interest [60]. In this case the power spectrum $|F(u,v)|^2$ or the autocorrelation function of the object can be estimated up to the diffraction limit of the telescope by dividing the short term power spectrum estimated from the object of interest by that of the unresolved point object.

It should be noted that the speckle interferometry gives information about the autocorrelation function only and not the desired object intensity. For the case of double star, it is possible to reconstruct the object because the object of interest in this case has a centre of symmetry [35]. However, for those objects without this property, losing phase information makes it difficult to reconstruct the object (Section 1.1).

There have been many algorithms being proposed to retrieve the lost phase information so that an object can be reconstructed by combining the magnitude information from speckle interferometry and the phase information from these algorithms. One technique is phase retrieval which reconstructs the Fourier phase from the Fourier magnitude uinsg constraints such as size and positivity. Alternative techniques such as Knox-Thompson or bispectrum reconstruct by further processing of the individual short term images.

3.2.1 Phase Retrieval

The original phase retrieval algorithm was proposed by Gerchberg and Saxton in 1972 [61]. It was intended for reconstructing the phase of the complex objects which ocurred in electron microscopy, and required knowledge of the magnitude of the image and its Fourier transform. It was modified for real objects by Fienup who substituted knowledge of the object's size and positivity for knowledge of its mangitude. The error reduction algorithm proposed by Fienup [62, 63] follows directly the philosophy of the Gerchberg and Saxton algorithm. It is based on the knowledge that the Fourier modulus of the object can be measured, the object is known to be non-negative and has a known support. The error reduction algorithm [62] can be summarized as follows,

1. form an initial estimate of the object denoted by $\hat{f}(x,y)$ and Fourier transform it to yield $\hat{F}(u,v)$.
2. extract the phase of $\hat{F}(u,v)$, denoted by $\theta(u,v)$, and combine it with the measured modulus in Fourier domain to form $F'(u,v)$, ie.,

$$F'(u,v) = |\hat{F}(u,v)|exp[i\theta(u,v)] \tag{3.13}$$

3. inverse Fourier transform $F'(u,v)$ to yield $f'(x,y)$.
4. a new function, $\hat{f}(x,y)$, is formed by incorporating known constraints in the image space domain, ie.,

$$\hat{f}(x,y) = \begin{cases} f'(x,y) & \text{if not violating any known constraints} \\ 0 & \text{if violating some known constraints.} \end{cases} \tag{3.14}$$

5. repeat Step 2 until convergence is achieved.

A common way to measure the convergence of the algorithm is to use the normalized root mean squared error metric which is defined in the Fourier domain as,

$$E_F = [\frac{\sum_u \sum_v [|\hat{F}(u,v)| - |F(u,v)|]^2}{\sum_u \sum_v |F(u,v)|^2}]^{1/2} \tag{3.15}$$

or in the image space domain as,

$$E_O = [\frac{\sum_{x,y\in\gamma} |f'(x,y)|^2}{\sum_{x,y} |f'(x,y)|^2}]^{1/2} \tag{3.16}$$

where γ denotes the space in which the known a priori constraint is violated. It has been shown that the error reduction algorithm converges in the weak sense that the normalized root mean squared error cannot increase with an increasing number of iterations. Some researchers found that although the normalized root mean squared error decreases rapidly for the first few iterations; it decreases extremely slowly for later iterations [63]. It may require an impractical large number of iterations for convergence which makes the algorithm unsatisfactory in practice.

In order to improve the rate of convergence of the error reduction algorithm, a modification version is used which is called the hybrid input-output algorithm [63]. The first three steps of the hybrid input-output algorithm are same as that of the error reduction algorithm. However, in the fourth step, it is modified as,

$$f(x,y) = \begin{cases} f'(x,y) & \text{if not violating any known constraints} \\ f(x,y) - \beta f'(x,y) & \text{if violating any known constraints} \end{cases} \tag{3.17}$$

where β is a constant feedback parameter. Notice that $f(x,y)$ is no longer an estimate of $f(x,y)$, it is instead an input function used to drive the output $f'(x,y)$ to satisfy the constraints. The hybrid input-output algorithm allows for considerable flexibility in the selection of the next input. However, there is no guarantee that the error will decrease on each iteration for the hybrid input output algorithm.

3.2.2 Knox-Thompson Approach

Knox-Thompson's (KT) algorithm [64–67] is analogous to speckle interferometry as it involves an ensemble averaging operation of many short-exposure images. This method, however, can retrieve the phase information about the object. In image space domain, the KT double correlation can be written as,

$$i^{KT}(x_1, \triangle u) = \int_{-\infty}^{\infty} i^*(x) i(x + x_1) exp[2\pi j \triangle u x] dx \tag{3.18}$$

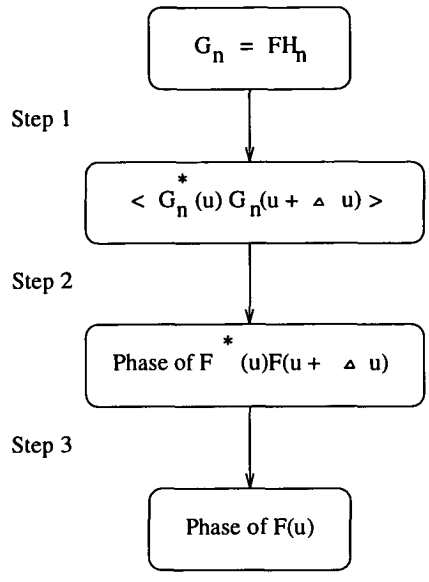


Figure 3.7: A block diagram illustrating the idea of the KT algorithm

where $i(x)$ is a two-dimensional intensity distribution function. Its cross power spectrum can be written as,

$$I^{KT}(u_1, \Delta u) = I(u_1)I^*(u_1 + \Delta u) \tag{3.19}$$

The KT algorithm is a modification of the autocorrelation technique. A Fourier component is multiplied with a component at a frequency displaced by a small vector Δu. The result is then averaged over an ensemble of speckles. At zero separation, $\Delta u = 0$, it is the auto-correlation function which does not contain any phase information. When $\Delta u \neq 0$, the cross-correlation function contains phase information about the object, provided the vector difference $-u - \Delta u$ is not outside the spatial frequency bandwidth [65].

A block diagram illustrating the idea of the KT algorithm is shown in Figure 3.7. Based on the mathematical model given in Eq (2.14) and are repeated here as,

$$G_n(u, v) = F(u, v)H_n(u, v) + C_n(u, v) \tag{3.20}$$

the object cross power spectrum is related to the image cross spectrum as,

$$< G_n{}^*(u)G_n(u+\Delta u) >= F^*(u)F(u+\Delta u) < H_n{}^*(u)H_n(u+\Delta u) > \qquad (3.21)$$

where $< H_n{}^*(u)H_n(u+\Delta u) >$ is called the cross spectrum transfer function. The phase relationship can be written as,

$$\begin{aligned} phase< G_n{}^*(u)G_n(u+\Delta u) > \; &= \; phaseF(u+\Delta u) - phaseF(u) \\ &\quad + phase< H_n{}^*(u)H_n(u+\Delta u) > \end{aligned} \qquad (3.22)$$

Eq (3.22) shows that the phase of the image cross power spectrum coincides with the object phase difference at frequencies $u+\Delta u$ and u, plus a term due to the cross spectrum transfer function. As was pointed out by Knox and Thompson [64], the effect of the cross spectrum transfer function can be compensated by calibrating the cross spectrum transfer function on a reference star. With this additional information, a ratio can be found which depends only on the object phases at points u and $u+\Delta u$, ie.,

$$\frac{< G_n{}^*(u)G_n(u+\Delta u) >}{| < G_n{}^*(u)G_n(u+\Delta u) > |} \; \frac{| < H_n{}^*(u)H_n(u+\Delta u) > |}{< H_n{}^*(u)H_n(u+\Delta u) >} =$$

$$\exp[j\ (phase(F(u+\Delta u)) - phase(F(u)))] \qquad (3.23)$$

provided that $\Delta u < r_0/\lambda$. When $\Delta u > r_0/\lambda$, the phase distribution of the term $H_n{}^*(u,v)H_n(u+\Delta u)$ is uniformly distributed between $-\pi$ and π. Thus the average of the term $F^*(u)F(u+\Delta u)$ becomes zero, ie.,

$$< F^*(u)F(u+\Delta u) >= 0 \qquad (3.24)$$

By ensuring that $\Delta u < r_0/\lambda$, the phase of $[F^*(u)F(u+\Delta u)]$ can be found up to the diffraction limit of the telescope. After finding the phase of the object cross spectrum, the problem remains to recover the phase of $F(u)$. A simple iterative technique is given as follows,

1. Set $u = 0$, $F^*(0)F(\Delta u)$ would give the value for different values of Δu

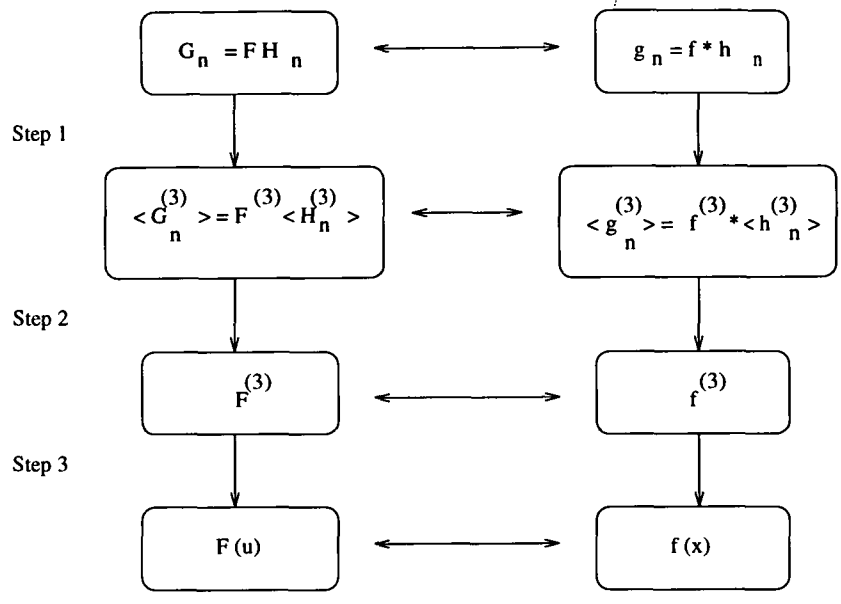


Figure 3.8: A block diagram showing the idea of the TC algorithm in both the space and the frequency domains.

2. Set $u = \triangle u$, $F^*(\triangle u)F(2\triangle u)$ will give another set of values.

Hence the phase of $F(u)$ can be built up iteratively. Because of this iterative scheme, the error in estimating $F(u)$ is accumulated which could be quite significant for large value of u.

3.2.3 Triple Correlation/Bispectrum Approach

Another method for retrieving the phase information about the object is by calculating the triple correlation function of the intensity distribution [65,67–73]. The triple correlation function is defined as,

$$i^{TC}(x_1, x_2) = \int_{-\infty}^{\infty} i^*(x)i(x + x_1)i(x + x_2)dx \tag{3.25}$$

and its Fourier transform pair, commonly called bispectrum, can be written as,

$$I^3(u_1, u_2) = I(u_1)I^*(u_1 + u_2)I(u_2) \tag{3.26}$$

Notice that the average bispectrum of the image is related to that of the object through the following equation,

$$< G_n^{(3)}(u_1, u_2) >= F^{(3)}(u_1, u_2) < H_n^{(3)}(u_1, u_2) > \tag{3.27}$$

The processing algorithm to recover the original object is shown in Figure 3.8. From Eq (3.27), $< H_n^{(3)}(u_1, u_2) >$ has to be estimated before $F^{(3)}(u_1, u_2)$ can be obtained. As was pointed out by Lohmann, et. al [68], $< H_n^{(3)}(u_1, u_2) >$ is real and non-zero over the diffraction limited portion of the telescope pupil. It follows that the phase of the image bispectrum must equal that of the object bispectrum. The problem is then reduced to reconstructing the phase of the object from the image bispectrum. To see how the phase of the object can be recovered from the image spectrum recursively, consider a one-dimensional sampled bispectrum $F_{l,m}^{(3)}$ which can be written as,

$$F_{l,m}^{(3)} = F_l F_m F_{-l-m} \tag{3.28}$$

Denoting the phase of the object spectrum as ϕ and that of the bispectrum as β, the phase relationship between the object and the bispectrum can be written as,

$$\phi_n = \phi_m + \phi_{n-m} - \beta_{n-m,m} \tag{3.29}$$

As the object is real and its absolute position is not of interest, ϕ_0, ϕ_1 are set to zero, and

$$\phi_n = \phi_1 + \phi_{n-1} - \beta_{n-1,1} \tag{3.30}$$

Hence, $\phi_2, \phi_3, \ldots$ can be found by simple recursion with $n = 2, 3, \ldots$. In fact, the process can be repeated for different values of m to yield some more estimates of the phase at each spatial frequency. A more sophisticated way to reconstruct the phase of the object from the image bispectrum is to employ a least square approach as described by Haniff [74].

3.3 Wavefront Sensing Deconvolution

The wavefront sensing deconvolution is a combination of the adaptive optics approach and the post-detection processing approach [51]. This technique requires a simultaneous

record of the atmospheric distortion information and the speckle image. In addition to the imaging camera that records an image of the object of interest, a wavefront sensor is needed to determine the wavefront at the same time when the image is recorded. This wavefront information can then be used to construct the instantaneous optical transfer function, which is the Fourier transform of the point spread function and thus providing a means of reconstructing the original object.

From the imaging camera, a set of turbulence degraded images denoted by $\{G_n(u,v)\}$ is obtained. Because of the random time variation of the atmospheric turbulence, each recorded speckle image is distorted differently and the lost frequencies are different for each speckle image. Hence the whole frequency spectrum of an object can in principle be reconstructed from a significantly large set of short exposure images provided that thier associated point spread functions are known.

The distorted wavefront information is usually provided by the wavefront sensor. The distorted wavefront and the optical transfer function are related as,

$$H_n(u,v) = e^{i\phi} \otimes e^{i\phi} \tag{3.31}$$

where $\otimes$ denotes correlation operator, and ϕ is the distorted wavefront phase. Eq (3.31) is obtained by assuming that the variation in the optical transfer function is a function of the phase variation in the aperture only. The effect of magnitude fluctuation induced by the atmosphere is neglected in this equation.

The problem which remains is to reconstruct the original object from $\{G_n(u,v)\}$ and $\{H_n(u,v)\}$. The processing algorithm can be summarized as shown in Figure 3.9. The object in Fourier domain is estimated by,

$$\hat{F}(u,v) = \frac{< G_n(u,v)H^*{}_n(u,v) >}{< |H_n(u,v)|^2 >} \tag{3.32}$$

It should be noted that

$$< |H_n(u,v)|^2 >= | < H_n(u,v) > |^2 + 0.342(\frac{r_0}{D})^2 T_0 \tag{3.33}$$

where $< H_n(u,v) >$ is the long term exposure optical transfer function and T_0 is the ideal optical transfer function of the telescope. The long term exposure optical transfer function can be constructed by summing all the short exposure transfer functions and

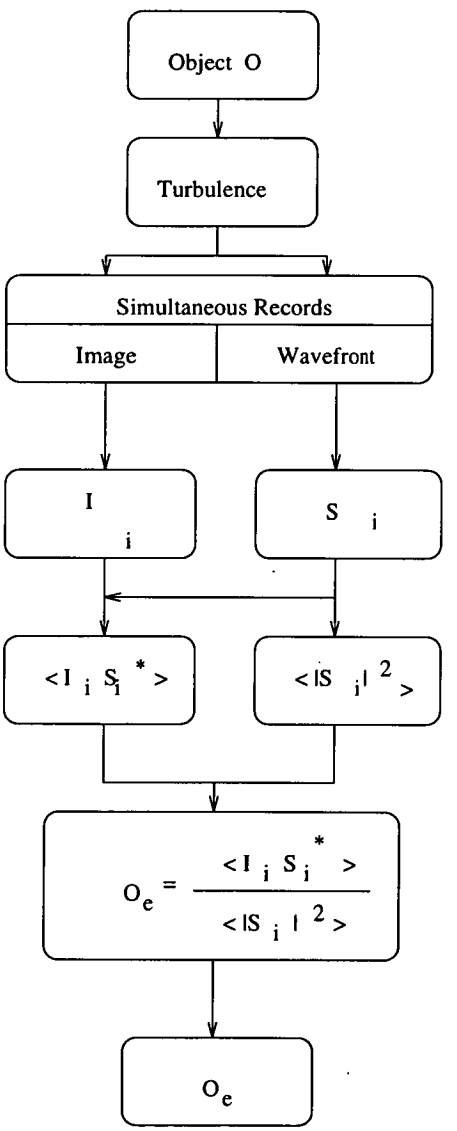


Figure 3.9: A block diagram showing the wavefront sensing deconvolution algorithm from the turbulence-degraded image to the object restoration.

then averaging them. As can be seen from Eq (3.33), the calculation of $\hat{F}(u,v)$ is always possible. The denominator is never equal to zero for frequencies less than the cut-off frequency which is the theoretical limit posed by diffraction. It can be shown that the estimated object, $\hat{F}(u,v)$, minimizes the mean squared error, E, which is defined as,

$$E =<< |\hat{F}(u,v)H_n(u,v) - G_n(u,v)|^2 >> \tag{3.34}$$

where the notation $<<>>$ represents the average over the space of all spatial frequencies.

This algorithm relies on the accuracy of the estimation of the wavefront as the optical transfer function depends on the distorted wavefront phase. In order to improve the estimation of the object spectrum, one has to estimate ϕ accurately. It is important to provide the best possible estimate of the wavefront for this technique to be effective, a point discussed further in Chapter 4.

3.4 Blind Deconvolution

All the post-detection processing techniques discussed so far require either some means to estimate the distortion introduced by the atmospheric turbulence or callibration by a reference star. In this section, we consider the case in which the distortion is considered to be an unknown. It is generally known as a blind deconvolution problem in which both the object and the point spread function need to be recovered [75]. Many of the arguments for the feasibility of blind deconvolution are identical to those developed to the phase retrieval problem, and indeed they both are only theoretically feasible in greater than one dimensional space [46].

Phase retrieval is a particular case of the blind deconvolution in which the point spread function can be considered to be equal to $f(-\mathbf{x})$. Thus a way to solve the blind deconvolution problem was in effect to combine two phase retrieval algorithmsin an iterative loop [44].Then by using some filtering techniques, $f(\mathbf{x})$ and $h(\mathbf{x})$ can be recovered iteratively.

An analytical method to solve the blind deconvolution problem is based on finding the zero-sheets of the Fourier transform of the object. It makes use of the fact that two

dimensional convolved image has distinct sets of zeros for the object and the point spread function in the absence of noise [46].

3.4.1 Zero Sheet Based Blind Deconvolution

Zero-sheet based blind deconvolution was proposed by Lane and Bates in 1987. This algorithm can be applied to deconvolve an image with multiple components of convolution, provided that the dimension of the image is greater than unity [46]. This algorithm is based on the fact that the analytic properties of the spectra of multi-dimensional images have compact support. Such spectra are necessarily zero on (2K-2) dimensional hyper-surfaces in a 2K dimensional space, which is constructed by treating each of the conventional real spectra co-ordinates as a complex variable.

For a two-dimensional image, two symbols, ζ and γ, are used to denote the complex variables. They are defined as,

$$\begin{aligned} \zeta &= \xi + j\eta = \exp(j2\pi u) \\ \gamma &= \alpha + j\beta = \exp(j2\pi v) \end{aligned} \tag{3.35}$$

where u and v are the Cartesian components of $\mathbf{u}$ in two-dimensional space, and ξ, η, α and β are all real. On representing a two-dimensional image as an $M \times M$ array of pixels, ie.,

$$\begin{aligned} f(\mathbf{x}) &= f(x,y) \\ &= \sum_{n=1}^{M-1} \sum_{m=0}^{M-1} f_{m,n}\delta(x - m\triangle x)\delta(y - n\triangle y) \end{aligned} \tag{3.36}$$

where $\delta(\cdot)$ denotes the Dirac delta function, $\triangle x$ and $\triangle y$ are the pixel spacings in the x and y directions respectively, and $f_{m,n}$ are the pixel values. Its spectrum can be written as,

$$\begin{aligned} F(\mathbf{u}) &= F(u,v) \\ &= F(\zeta,\gamma) \\ &= \sum_{n=1}^{M-1} \sum_{m=0}^{M-1} f_{m,n}\zeta^m\gamma^n \end{aligned} \tag{3.37}$$

The finite polynomial formed defined by Eq (3.37) has the advantage that the spectrum is completely represented by a finite number of zeros. Eq (3.37) represents a four dimensional problem. The zeros form a continuous sheet (two-dimensional surfaces) in ξ, η, α and β space. The computation of a two-dimensional zero sheet is done by initially fixing one of the complex variables in Eq (3.37), say γ. The point zeros of the resultant polynomial in the other complex variable ζ are then evaluated. The next step is to increment γ successively and re-evaluate the point zeros in ζ, each of which has been shifted by only a small amount. This incremental shifting makes the process of zero location more efficient since a given set of point zeros provides good initial estimates for the next set of zeros.

By repeatedly incrementing γ, one forces each point zero to track across a portion of the zero sheet, until the complete zero sheet is eventually traced out. The increments in γ are small enough that the displacement of each point zero is unambiguously determined from one increment to the next. So each zero sheet, being necessarily continuous, is unambiguously and separately mapped out. Then each Z spectrum is separately reconstructed, thereby permitting each image component to be computed separately.

In the absence of noise, the zero-sheet based blind deconvolution algorithm provides a good estimate of the components of the convolution as was demonstrated by Lane and Bates [46]. However, all real world data are corrupted by some form of noise. In this case, the zero sheet is similar to the union of all the zero sheets of the components of convolution, but with the latter being interconnected by 'bridges' [76]. As the noise level increases, the bridges become more numerous and more prominent. As a result, the zero sheets composing the original convolution are unrecognizable for high level of noise contamination. A practical technique for dealing with the levels of noise remains to be found.

3.4.2 Iterative Blind Deconvolution Loop

The iterative blind deconvolution loop was first proposed by Ayers and Dainty in 1988 [44]. It basically combines two phase retrieval algorithms and it imposes some a priori information concerning the object and the point spread function onto the estimation process. A block diagram demonstrating the idea of the iterative blind deconvolution loop is shown in Figure 3.10.

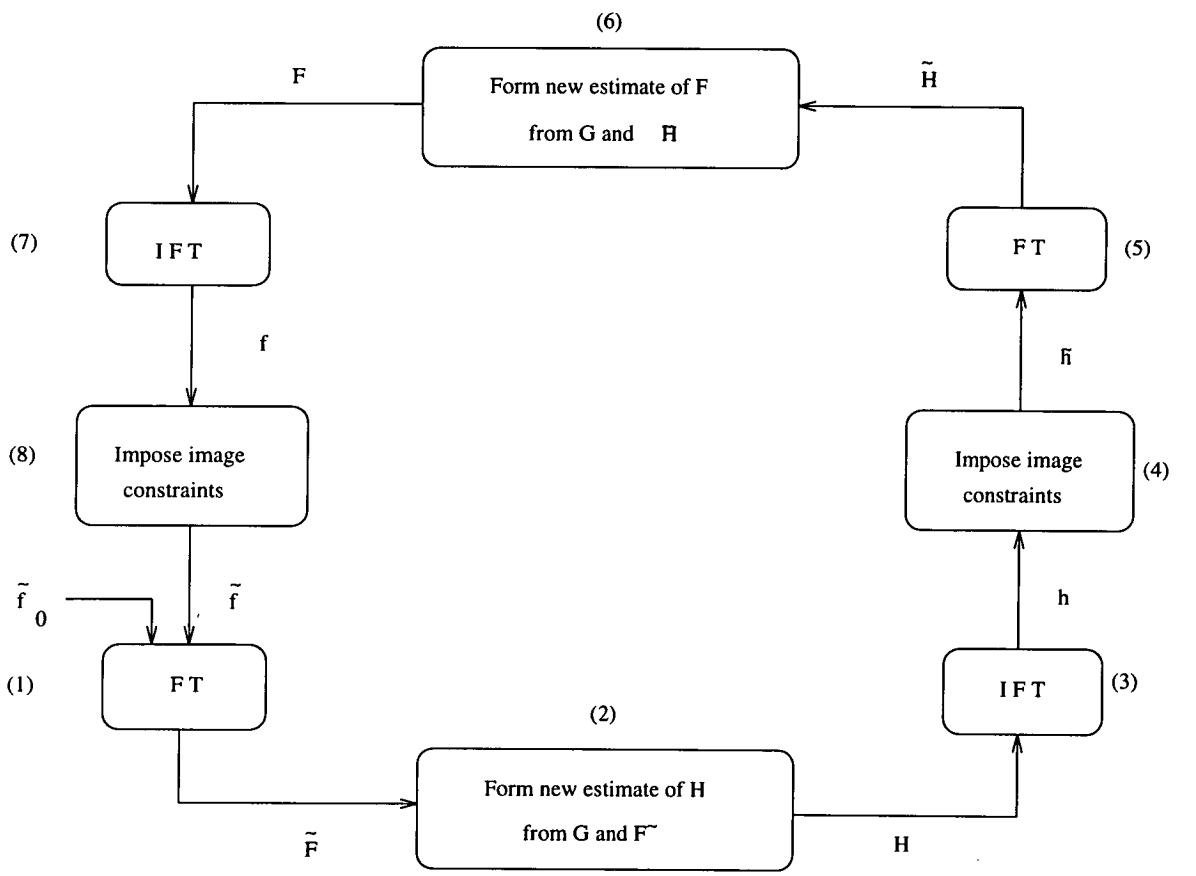


Figure 3.10: A block diagram illustrating the steps involved in the iterative blind deconvolution loop.

First a non-negative initial estimate, $\tilde{f}_0$, is input into the iterative scheme. This function is Fourier transformed to yield $\tilde{F}_0(\mathbf{u})$ (Step 1). The estimated Fourier spectrum is then combined with the known convolution $G(\mathbf{u})$ to form the first estimate of the spectrum of the psf $H_0(\mathbf{u})$ (Step 2). This can be done by using Wiener filtering or some other filtering techniques. Inverse Fourier transforming $H_0(\mathbf{u})$ then gives $h_0(\mathbf{x})$ (Step 3).

The image domain constraints, such as non-negativity constraint and the support constraint, are now imposed to $h_0(\mathbf{x})$ (Step 4). This can be done by putting to zero all points of the point spread function $h_0(\mathbf{x})$ that violate its image domain constraints. A positive constrained estimate $\tilde{h}_0(\mathbf{x})$ is consequently formed which is Fourier transformed to give the spectrum $\tilde{H}_0(\mathbf{u})$ (Step 5).

This is then used to generate the next spectrum estimate $F_1(\mathbf{u})$ (Step 6) by using appropriate filtering techniques which is then inverse Fourier transformed to give $f_1(\mathbf{x})$ (Step 7). A single iterative loop is completed by constraining this function to satisfy its image domain constraint which yields the next function estimate $\tilde{f}_1(\mathbf{x})$ (Step 8). The iteration loop is repeated until the convolution of the estimated object and the estimated point spread function give the required convolution, $g(\mathbf{x})$, to the required accuracy.

This algorithm is conceptually simple. However, there are some difficulties associated with this algorithm. One difficulty with the algorithm is in the regions where $G(\mathbf{u})$ is small. As

$$\frac{\partial F(\mathbf{u})}{\partial G(\mathbf{u})} = \frac{1}{H(\mathbf{u})} \tag{3.38}$$

small errors in $G(\mathbf{u})$ can result in large errors in $F(\mathbf{u})$ in the vicinity of the zeros of $H(\mathbf{u})$. Hence the algorithm is prone to large errors in reconstruction. In order to solve this problem, the following scheme is employed [77]. If $|G(\mathbf{u})| <$ noise level, then

$$F_{i+1}(\mathbf{u}) = \tilde{F}_{\mathbf{i}}(\mathbf{u}) \tag{3.39}$$

If $|\tilde{H}_i(\mathbf{u})| \geq |G(\mathbf{u})|$,

$$F_{i+1}(\mathbf{u}) = (1-\beta)\tilde{F}_{\mathbf{i}}(\mathbf{u}) + \beta\frac{G(\mathbf{u})}{\tilde{H}_{\mathbf{i}}(\mathbf{u})} \tag{3.40}$$

If $|\tilde{H}_i(\mathbf{u})| < |G(\mathbf{u})|$,

$$\frac{1}{F_{i+1}(\mathbf{u})} = \frac{1-\beta}{\tilde{F}_i(\mathbf{u})} + \beta\frac{\tilde{H}_i(\mathbf{u})}{G(\mathbf{u})} \tag{3.41}$$

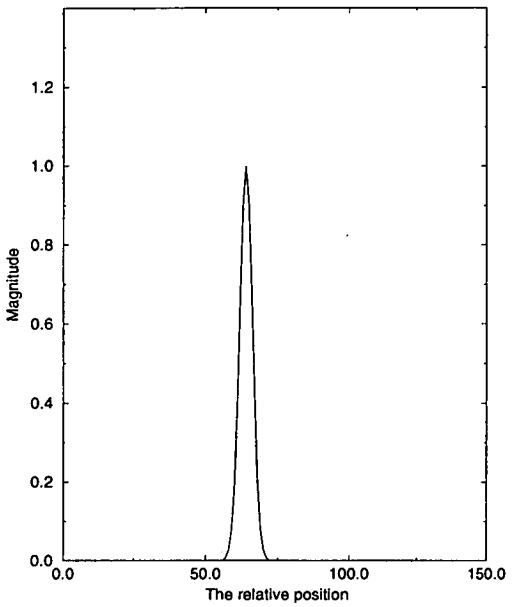


Figure 3.11: A Gaussian signal used to demonstrate the effect of the noise in different frequency bands.

for β between 0 and 1.

Another difficulty is that it cannot be equated to a steepest descent search on an error metric in either the Fourier or image space, where the error metric is defined as a quantitative measure of how much the current estimate violates the known constraints [78]. As a result, performing an extra iteration of the algorithm may result in new estimates of $f(\mathbf{x})$ and $h(\mathbf{x})$ which are worse than the estimates before the iteration.

However, in the presence of noise, this algorithm has an important advantage over other blind deconvolution algorithms that do not use any type of filtering in the estimation process. In order to demonstrate this point, consider the following example. The object is shown in Figure 3.11 which is a Gaussian signal. For reference, a uniformly distributed random noise is generated with a signal to noise ratio of 0 dB. Their magnitude spectra are shown in Figure 3.12. The object spectrum has some large values in the low frequency range and decays quickly as the frequency increases. In contrast, the noise spectrum is quite uniform in the whole frequency range. As a result, the signal to noise ratio in the low frequency range is higher than that in the high frequency range. And the contribution from the signal and that from the noise cannot be distinguished in the high frequency part.

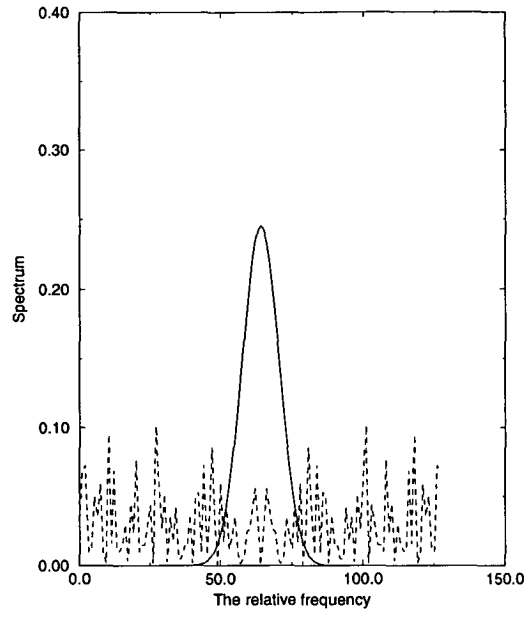


Figure 3.12: A plot of the relative magnitude of the object and the noise spectra versus the frequency. The solid line is the object spectrum while the dotted line is the noise spectrum

When the iterative loop is applied with an appropriate choice of cutoff frequency, the algorithm can effectively suppress the effect of the noise in the high frequency range, and thus prevent them from polluting the reconstruction of the object. For those blind deconvolution algorithms without using any filtering techniques, the low SNR in the high frequency range can make the reconstruction extremely noisy. This problem of estimating a suitable cutoff is one of superresolution and is discussed further in Chapter 7 [79].

Chapter 4

Wavefront Estimation

Both the real time adaptive optics technique and the computer post-detection processing techniques require an estimate of the wavefront distortion information before these techniques can be applied to reconstruct the diffraction-limited object. Indeed, the wavefront estimation plays a vital role in compensating the disturbances associated with the atmospheric turbulence.

The wavefront estimation problem can basically be rewritten in matrix notation as (Section 3.1),

$$\mathbf{h} = \Theta\mathbf{a} \tag{4.1}$$

This chapter discusses the solution to this equation. Firstly, the problem is formulated mathematically in Section 4.1. In particular, two basis functions that are commonly used to model the wavefront distortion, namely the Zernike polynomials and the Karhunen-Loeve polynomials, are discussed.

A simple and straightforward solution to Eq (4.1) is the pseudo-inverse solution. It should be noted that the number of measurements from the Shack Hartmann sensor is finite, but the degree of freedom of the wavefront is infinite which also implies that the number of unknowns needs to be estimated is infinite. It is a fundamentally ill-conditioned problem. In practice the pseudo-inverse method can only be used to solve for the number of unknowns that are smaller than the number of measurements available. The pseudo-inverse solution together with simulation results are presented in

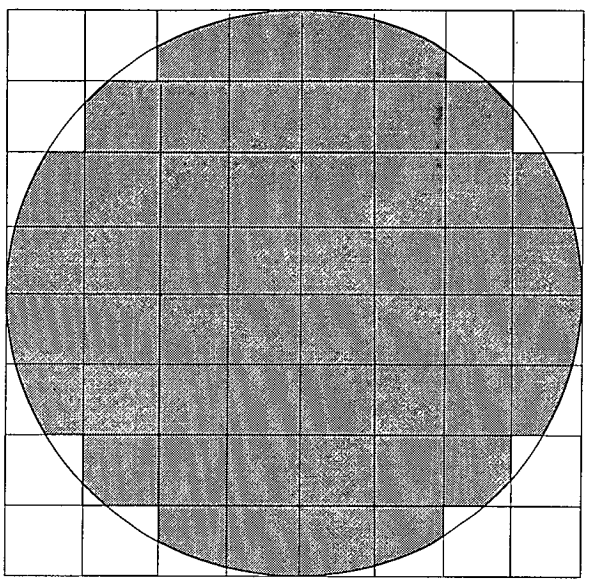


Figure 4.1: The 8 x 8 Shack-Hartmann array of lenslets superimposed on a circular aperture. The shading is used to indicate those lenslets whose outputs are used in the calculation of the coefficients of the basis functions.

Section 4.2.

In order to solve the under-determined problem, some extra constraints need to be applied to ensure the estimation problem becomes well-conditioned, ie., the number of unknowns is always less than or equal to the number of 'measurements' available. It is the principle of the weighted least square approach and is discussed in Section 4.3. The weighted least square approach can be shown to be same as the estimations proposed by Wallner [80] and the MAP solutions proposed by Solomon et. al [81]. They are covered in Section 4.4 and Section 4.5 respectively.

4.1 Problem Formulation

A Shack-Hartmann sensor gives two measurements in two orthogonal directions for each sub-aperture as discussed in Section 3.1. Figure 4.1 shows a 8×8 lenslet array superimposed on an unobscure circular telescope aperture. It is readily apparent that not all sub-apertures are equally illuminated which means that the accuracy of the measurements in different illuminated sub-apertures are different. There are a number

of possibilities for overcoming the differences in illumination of the arrays. The simplest approach, that of ignoring all sensor measurements obtained from sub-apertures which are not fully illuminated, wastes a significant amount of data.

In this analysis, only those sub-apertures of the Shack-Hartmann sensor with more than half-illumination are considered in order to eliminate apertures that would cause difficulties in low light level. It should be noted that this choice of threshold is somewhat arbitrary and in reality would depend on a number of practical factors. Let N be the total number of sub-apertures that satisfy this condition. Thus in Figure 4.1, the effective number of sub-apertures is 52 and the total number of available measurements is 104, since each sub-aperture gives two measurements [82].

The incident wavefront can be described in terms of the Cartesian co-ordinates (x, y) located in the aperture of the telescope. As mentioned in Section 3.1, there are two approaches to modelling the incoming wavefront, zonal or modal. As shown by Southwell [56], the modal approach is better than the zonal approach since the reconstructed wavefront is less sensitive to the measurement noise, the computation time is shorter and the number of modes considered can be adapted to a particular problem. Consequently only the modal approach is considered here. The modal expansion of the incident wavefront onto the telescope aperture can be written in terms of a set of orthogonal basis functions as,

$$\phi(x, y) = \sum_{i=0}^{\infty} a_i \psi_i(x, y) \tag{4.2}$$

where $\phi(x, y)$ represents the unknown wavefront incident onto the telescope aperture, $\psi_i(x, y)$ is the i^{th} basis function, and a_i is the weighting attached to the i^{th} basis function.

The behaviour of the atmospheric turbulence is known to follow a Kolmogorov law (Section 2.2) and, as a result, its statistical behaviour is well known [30]. The optimal set of basis functions to describe a wavefront aberrated by Kolmogorov turbulence are the Karhunen-Loeve functions [83, 84]. Choosing $\psi_i(x, y)$ to be Karhunen-Loeve functions means the coefficients a in Equation (4.2) are statistically independent. Unfortunately, it is not possible to calculate these functions analytically. However, Wang and Markay [85] noted that for a circular aperture the low order Karhunen-Loeve functions are closely approximated by the Zernike polynomials. Hence, Zernike polynomials

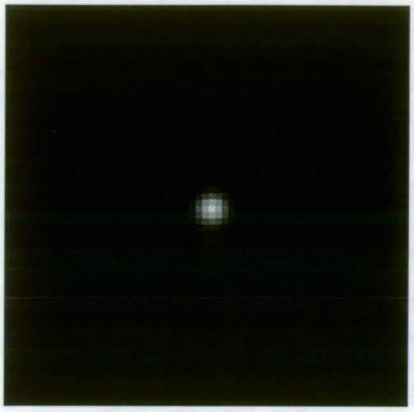


Figure 4.2: An image of a point source object which is formed with no atmospheric turbulence and observed through a circular aperture.

are often used as the basis functions.

It should be noted that as shown in Table 1.1, $Z_1(x, y) = 1$, a constant over the aperture. For a pure Kolmogorov turbulence, the weighting coefficient associated with this function is infinite. Fortunately, the average phase across the aperture does not affect the image and thus can be left out of the basis function, ie.,

$$\phi(x, y) = \sum_{n=2}^{\infty} a_n Z_n(x, y) \tag{4.3}$$

The following example shows the effect of the aberrations in Equation (4.3). Figure 4.2 shows the object observed when there is no aberration. Figure 4.3 is a typical image obtained when observed through the turbulent atmosphere with $D/r0 = 10$. Figure 4.4 shows the image with tip/tilt (ie., Z_2 and Z_3)removed. Figure 4.5 shows the image with the first 10 orders of the Zernike polynomials compensated. Figure 4.6 shows the image with the first 100 orders of the Zernike polynomials compensated. As can be seen, as more orders of the Zernike polynomials are compensated, the resultant image approaches the ideal case where there is no aberration. In this simulation, we can also see that the tip/tilt contribute a large proportion of the error associated with the

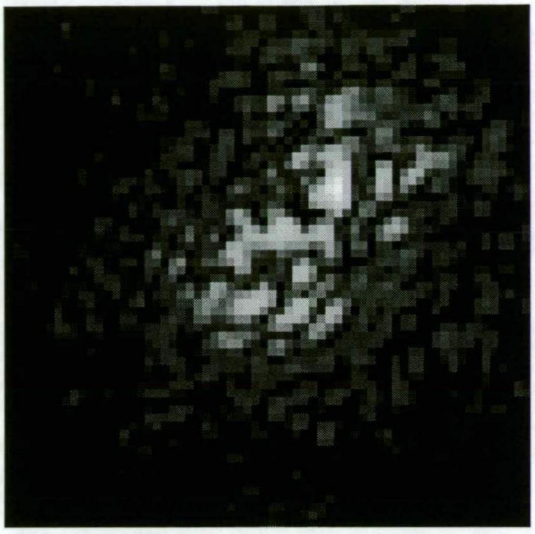


Figure 4.3: A typical image obtained when observed through the turbulent atmosphere.

Figure 4.4: An image formed with the first two coefficients of the Zernike polynomial compensated.

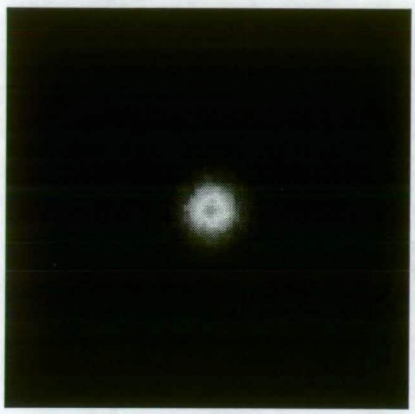


Figure 4.5: An image formed with the first ten coefficients of the Zernike polynomial compensated.

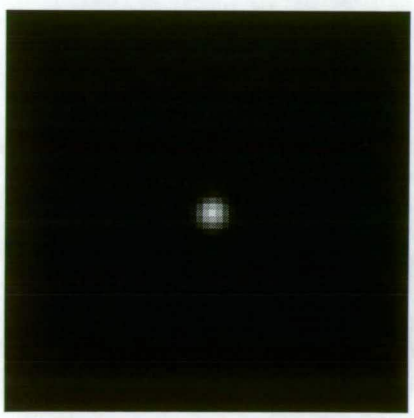


Figure 4.6: An image formed with the first hundred coefficients of the Zernike polynomial compensated.

aberration. With this removed, there is a significant improvement in the reconstruction of the wavefront.

4.2 General Least Square Approach

In practice, $\mathbf{\Theta}$ and $\mathbf{H}$ are available and the problem of estimating the wavefronts thus reduces to an algebraic problem, namely solving a system of linear equations. The incident wavefront is expanded in terms of a set of infinite number of Zernike polynomials. However, there are only finite number of measurements from the Shack-Hartmann sensor, it is also an under-determined problem. It is well known that the number of coefficients that can be estimated must be smaller than the number of measurements available.

The first approach to estimate these coefficients is thus to estimate only the first X coefficients in the expansion, where $X \leq 2N$. All other coefficients are then put to zero, ie.,

$$\hat{\mathbf{a}} = (\mathbf{\Theta^T \Theta})^{-1} \mathbf{\Theta^T h} \tag{4.4}$$

where the symbol $\hat{\mathbf{a}}$ indicates an estimated quantity of $\mathbf{a}$. The mean square error in this estimation can be computed by,

$$E = < (\phi(x,y) - \hat{\phi}(x,y))^2 > \tag{4.5}$$

where $\hat{\phi}(x,y)$ is the reconstructed wavefront by using $\hat{\mathbf{a}}$. It is more convenient, however, to employ Parseval's theorem and compute the error using the following form,

$$E = < \sum_{n=2}^{X+1} (a_n - \hat{a}_n)^2 > + < \sum_{n=X+2}^{\infty} a_n^2 > \tag{4.6}$$

The residual error, E, consists of two parts. The first part is the error in estimating the first X coefficients. The second part is the error resulting not estimating any coefficients of an order higher than X.

The simulation results are shown in Figure 4.7, Figure 4.8 and Figure 4.9 for 2×2, 4×4 and 8×8 Shack-Hartmann sensor respectively. They provide 8, 24 and 104 measurements respectively. As shown from these curves, there is an initial steep decline

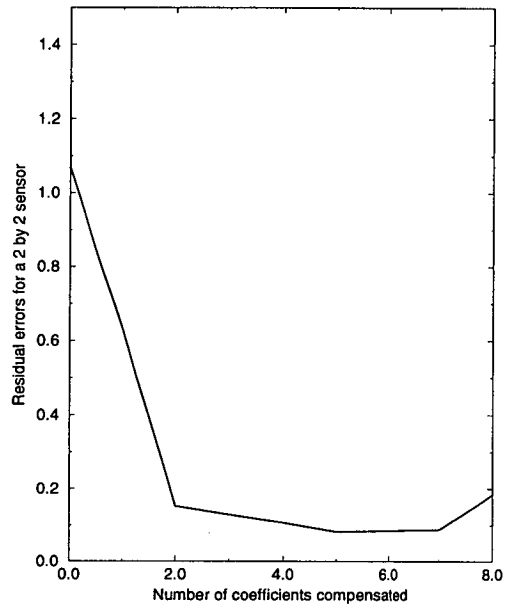


Figure 4.7: The residual errors for a 2 × 2 Shack Hartmann sensor using the general least square approach.

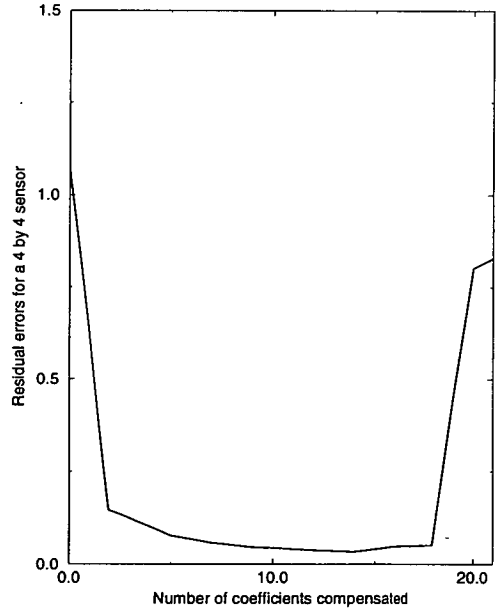


Figure 4.8: The residual errors for a 4 × 4 Shack Hartmann sensor using the general least square approach.

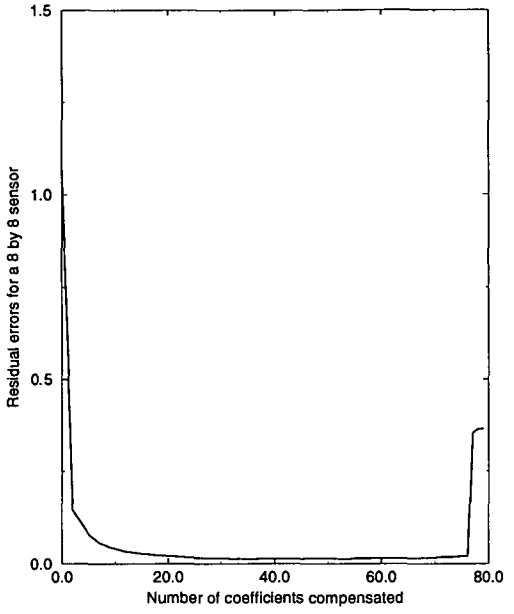


Figure 4.9: The residual errors for a 8 × 8 Shack Hartmann sensor using the general least square approach.

in the residual error. But after passing the minimum point, estimating more coefficients no longer results in any improvement in the estimation. In fact, the residual error starts to increase when the number of coefficients estimated gets close to the number of measurements available.

One of the reasons for the increase in the residual errors in Figure 4.7, Figure 4.8 and Figure 4.9 is due to the partial illumination of the sub-apertures. As can be seen in Figure 4.1, not all the sub-apertures have the same illumination. Hence the accuracy of the measurements from each sub-aperture may be different and thus the signal to noise ratio is different from one sub-aperture to another. A more general model should be written as,

$$\mathbf{h} + \mathbf{v} = \Theta\mathbf{a} \tag{4.7}$$

where $\mathbf{v}$ models the noise term. In order to remove the effect of partial illumination, $\hat{\mathbf{a}}$ can be estimated by using the least square method weighted by different noise level of each sub-aperture. In this case, $\hat{\mathbf{a}}$ is estimated from the following equation,

$$\hat{\mathbf{a}} = (\Theta^{\mathbf{T}}\mathbf{D}\Theta)^{-1}\Theta^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{h} \tag{4.8}$$

where $\mathbf{D}$ is the covariance matrix of the noise level of each sub-aperture. It is a function

of the number of photons and the CCD arrangement [51,86]. The noise can be expressed as either,

1. in the case of photon-noise limitation,

$$\sigma_M = \frac{\pi}{\sqrt{2n_a}} \frac{\delta_a d'}{\lambda f} \tag{4.9}$$

where n_a is the mean number of detected photons within a sub-aperture, δ_a is the image size in the lenslet focal plane, and d' is the lenslet diameter.

2. in the case of diffraction-limited image and electronic noise,

$$\sigma_M = \frac{4}{\sqrt{3}} \frac{\sigma_n}{G_s N_s \epsilon^3} \tag{4.10}$$

where ϵ is the ratio between the spot width and the sub-aperture width in the CCD plane, G_s is the maximum signal in an image, σ_n is the root mean square error per pixel induced by the electronic noise, and N_s is the number of pixels in one direction used for the centroiding.

However, even using the weighted estimates by considering the different noise level of each sub-aperture, it is apparent that for a 2×2 sensor, each sub-aperture has the same level of illumination. This method thus produces an improvement only when there is sensor larger than 2×2.

Another reason for the increase in the residual errors is that the residual errors due to the higher order modes are ignored, ie., only the first X coefficients are estimated in both Eq (4,4) and Eq (4.8). According to conventional least square theory [87], unless the theoretical residuals are uncorrelated with each other, the estimates obtained will be biased and sub-optimal. Since there is a strong correlation between the measurements taken from the sub-apertures of the sensor, we cannot expect to obtain an optimal estimate.

4.3 Weighted Least Square Approach

The under-determined problem can be solved by incorporating some of the a priori knowledge into the estimation process so that the problem becomes well-conditioned.

One a priori information that is available is the covariance matrix of the weightings a_j. By assuming Kolmogorov turbulence, the covariance matrix of the weightings of Zernike polynomials can be derived theoretically [7]. If this piece of information is incorporated, a better estimation result is expected.

The a priori knowledge about the weightings a_j can be incorporated by adding one additional equation for each unknown, namely the knowledge that the expected value of this coefficient is zero. This leads the estimation to be over-estimated, ie., the number of equations must exceed the number of unknowns. The problem can then be reformulated as follows,

$$\mathbf{h}' + \mathbf{v}' = \mathbf{\Theta}'\mathbf{a} \tag{4.11}$$

where

$$\mathbf{\Theta}' = \begin{pmatrix} \mathbf{\Theta} \\ \mathbf{I} \end{pmatrix} \tag{4.12}$$

$$\mathbf{h}' = \begin{pmatrix} \mathbf{h} \\ \mathbf{0} \end{pmatrix} \tag{4.13}$$

and

$$\mathbf{v}' = \begin{pmatrix} \mathbf{v} \\ \mathbf{a} \end{pmatrix} \tag{4.14}$$

$\mathbf{I}$ and $\mathbf{0}$ are the identity matrix and the zero matrix respectively. Since every unknown contributes an additional equation, we effectively increase the dimensions of vectors $\mathbf{\Theta}$, $\mathbf{h}$ and $\mathbf{v}$ such that it is no longer an under-determined problem. By the conventional weighted least square approach [87], the weightings of the Zernike polynomials can be obtained by,

$$\hat{\mathbf{a}} = (\mathbf{\Theta}'^{\mathbf{T}}\mathbf{R}\mathbf{\Theta}')^{-1}\mathbf{\Theta}'^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{h} \tag{4.15}$$

where

$$\begin{aligned} \mathbf{R} &= \mathbf{E}[\mathbf{v}'\mathbf{v}'^{\mathbf{T}}] \\ &= \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} \end{aligned} \tag{4.16}$$

and

$$\mathbf{C} = \mathbf{E}[\mathbf{a}\mathbf{a}^{\mathbf{T}}] \tag{4.17}$$

The estimated weighting of the Zernike polynomials can then be written as,

$$\hat{\mathbf{a}} = (\Theta^{\mathbf{T}}\mathbf{D}^{-1}\Theta + \mathbf{C}^{-1})^{-1}\Theta^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{h} \tag{4.18}$$

As mentioned before, $\mathbf{C}$ is the covariance of the coefficients of the Zernike polynomials defined by

$$\mathbf{C} = \begin{pmatrix} E[a_2a_2] & E[a_2a_3] & \ldots & E[a_2a_p] \\ E[a_3a_2] & E[a_3a_3] & \ldots & E[a_3a_p] \\ \cdot & \cdot & \cdot & \cdot \\ E[a_pa_2] & E[a_pa_3] & \ldots & E[a_pa_p] \end{pmatrix} \tag{4.19}$$

It was pointed out by Noll [7] that the coefficients a_j can be considered to be Gaussian random variables with zero mean. For the Kolmogorov spectrum, Noll readily derived an expression for $< a_i a_j >$ which formed an element of the covariance matrix $\mathbf{C}$. The matrix element of $\mathbf{C}$ is given by,

$$\begin{aligned} < a_i a_j > \; = \; & \frac{0.046}{\pi}\frac{D}{r_0}^{5/3}[(n+1)(n'+1)]^{1/2}(-1)^{(n+n'-2m)/2}\delta_{mm'}\delta_{ll'} \\ & \int_0^\infty k^{-8/3}\frac{J_{n+1}(2\pi k)J_{n'+1}(2\pi k)}{k^2}dk \end{aligned} \tag{4.20}$$

where $J_n(z)$ is a Bessel function. The first few matrix elements of C are shown in Table 4.1.

It should be noted that $\mathbf{C}$ has infinite dimensions. In order to overcome the problem of inverting an infinite matrix, we again estimate only the first P coefficients of the Zernike polynomials as before. In contrast to the previous technique, however, we treat the remaining unestimated terms as noise and calculate numerically this contribution to the covariance matrix $\mathbf{D}$. Thus all modes are either estimated or their effects are accounted for in the expected residual noise covariance matrix. We can then allow P to increase past the number of sensor measurements and see its effect on the estimation. The results are shown in the Figure 4.10, Figure 4.11 and Figure 4.12 which examine the performance of 2×2, 4×4 and 8×8 Shack-Hartmann sensors respectively.

As P increases, the error in the wavefront estimation decreases monotonically for all three cases. The errors in the wavefront estimation achieved in the case are significantly lower than that achieved by the simple pseudo-inverse method as shown in Table 4.2.

	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
Z_2	0.4557	0	0	0	0	0	-0.0144	0	0
Z_3	0	0.4557	0	0	0	-0.0144	0	0	0
Z_4	0	0	0.0236	0	0	0	0	0	0
Z_5	0	0	0	0.0236	0	0	0	0	0
Z_6	0	0	0	0	0.0236	0	0	0	0
Z_7	0	-0.0144	0	0	0	0.0063	0	0	0
Z_8	-0.0144	0	0	0	0	0	0.0063	0	0
Z_9	0	0	0	0	0	0	0	0.0063	0
Z_{10}	0	0	0	0	0	0	0	0	0.0063
Z_{11}	0	0	-0.0039	0	0	0	0	0	0
Z_{12}	0	0	0	0	-0.0039	0	0	0	0
Z_{13}	0	0	0	-0.0039	0	0	0	0	0

Table 4.1: The first few matrix elements of $\mathbf{C}$

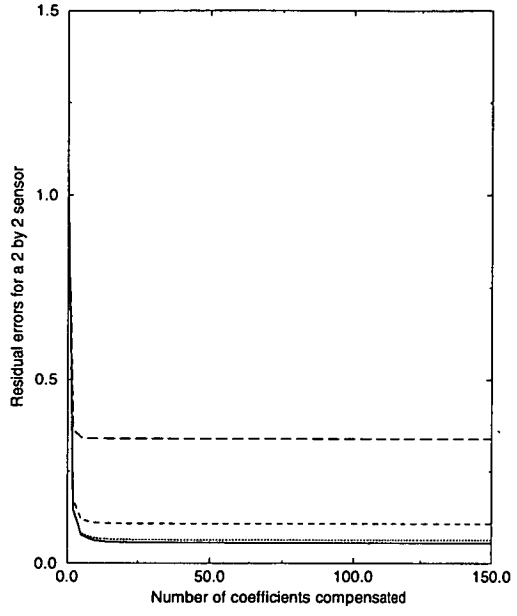


Figure 4.10: The residual errors for a 2 × 2 Shack Hartmann sensor using the weighted least square approach in the presence of different noise level. The solid line, the dotted line, the dashed line and the long dashed line represent SNR of ∞, 20 db, 10 db and 0 db respectively.

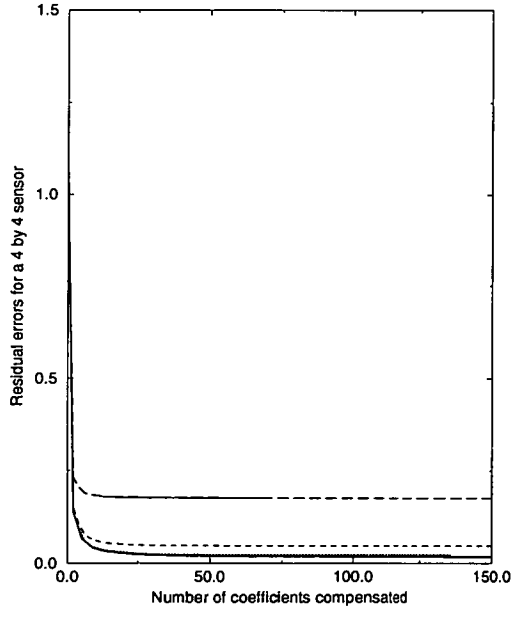


Figure 4.11: The residual errors for a 4 × 4 Shack Hartmann sensor using the weighted least square approach in the presence of different noise level. The solid line, the dotted line, the dashed line and the long dashed line represent SNR of ∞, 20 db, 10 db and 0 db respectively.

Sensors' size	SNR	minimum mean square errors
2 × 2	0 db	0.338
	10 db	0.107
	20 db	0.063
	∞	0.056
4 × 4	0 db	0.176
	10 db	0.048
	20 db	0.023
	∞	0.018
8 × 8	0 db	0.107
	10 db	0.042
	20 db	0.019
	∞	0.006

Table 4.2: The minimum mean square residual errors achieved by using the weighted least square approach for different size of the Shack-Hartmann sensor at different noise level.

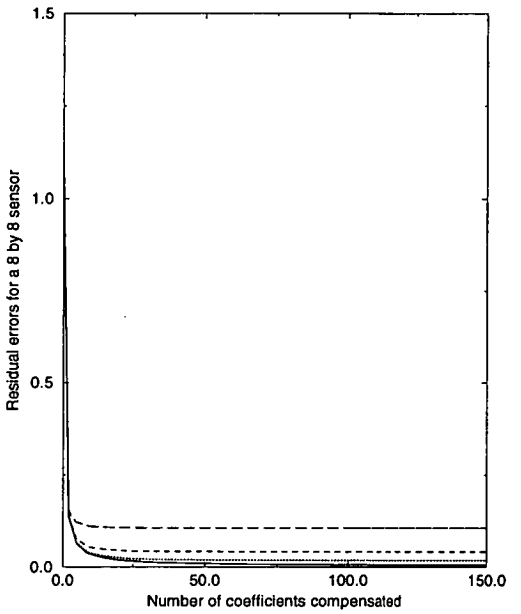


Figure 4.12: The residual errors for a 8×8 Shack Hartmann sensor using the weighted least square approach in the presence of different noise level. The solid line, the dotted line, the dashed line and the long dashed line represent SNR of ∞, 20 db, 10 db and 0 db respectively.

In practice, there is a point where the contribution of the higher order modes in the wavefront to the Shack-Hartmann measurements is much less than the errors in the sensor measurements. This is apparent in the diagrams shown in Figure 4.10, Figure 4.11 and Figure 4.12 . We can see that the improved performance obtained by increasing the number of modes estimated in the presence of noise flattens out rapidly and there is a point above which no obvious improvement can be made by estimating any more higher order modes. There is not the degradation observed in the general least squares approach of Section 4.2.

It should be noted that by incorporating the effects of unestimated mode in the residual error, we have introduced some correlation between the errors and the estimated coefficients, which we have assumed no correlation in the derivation. In practice, with the presence of noise, as P increases, we eventually reach a point where the sensor measurement noise dominates these correlations and hence they can be neglected. Also, as in the no noise case, the residual error obtained is monotonically decreasing. It shows that the performance is better if we estimate the higher order modes using a priori information rather than ignoring them.

4.4 Analysis of the Technique Proposed By Wallner

In the section, the result given by Wallner [80] is summarized here but with the notations we defined above. The overall correction system is represented as a number of sensors that measure wavefront slope, a corrector with a number of actuators, and a control law connecting sensors to actuators. All these components are interdependent, so a unified treatment of the overall system is essentail. The sensor measurement can be written as (cf Eq (4.7)),

$$s_n = \sum_{l=1}^{\infty} \theta_{nl} a_l + v_n \tag{4.21}$$

The control law for the system generates a command to each actuator of the wavefront corrector based on all the sensor outputs. For a linear control law, the command to the j^{th} actuator is defined as,

$$c_j = \sum_n M_{jn} s_n \tag{4.22}$$

where c_j is the command to the j^{th} actuator and M_{jn} is the weighting of n^{th} sensor signal in j^{th} actuator command. The responses of the actuators are combined to form the total wavefront correction as,

$$\hat{\phi}(x,y) = \sum_j c_j r_j(x,y) \tag{4.23}$$

where $\hat{\phi}(x,y)$ is the total wavefront correction and $r_j(x,y)$ is the response of the j^{th} actuator to a unit command.

The residual wavefront error can be expressed as,

$$\begin{aligned} \epsilon(x,y) &= \hat{\phi}(x,y) - \phi(x,y) \\ &= \sum_j r_j(x,y) \sum_n M_{jn} s_n - \phi(x,y) \end{aligned} \tag{4.24}$$

$$\tag{4.25}$$

The overall measure of performance is the expected mean square error across the aperture and is defined as,

$$< {\epsilon_c}^2 >= \int_A < \epsilon^2(x,y) > dxdy \tag{4.26}$$

Wallner proposed to minimize $< \epsilon_c^2 >$ so that the system performance is optimized. Thus by differenting $< \epsilon_c^2 >$ with respect to element of M_{jn} and equating to zero yields

the optimal estimator as,

$$R_{ji}^{-1} A_{in} S_{np}^{-1} \tag{4.27}$$

where S_{np} is the integral involving the product of two sensor functions, A_{in} is the integral involving the products of sensor and the actuator response function, and R_{ji} is the integral involving the products of two actuator response functions. These three terms are defined mathematically as,

$$\begin{aligned} R_{ji} &= \int_A r_j(x,y) r_i(x,y) dxdy \\ A_{in} &= \int_A r_i(x,y) < s_n \phi(x,y) > dxdy \\ S_{np} &= < s_n s_p > \end{aligned} \tag{4.28}$$

(4.29)

where $\int_A$ denotes integrating over the aperture.

As Zernike polynomial is a set of orthonormal polnomials in a circular aperture (Chapter 1), then R_{ji} can be written using the matrix notation as,

$$\mathbf{R} = \mathbf{I} \tag{4.30}$$

The second term in Eq (4.27) can be simplied as,

$$\begin{aligned} A_{in} &= \int_A r_i(x,y) < \sum_{l=1}^{\infty} \theta_{nl} a_l \sum_{k=1}^{\infty} a_k r_k(x,y) > dxdy \\ &= \int_A r_i(x,y) \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} \theta_{nl} < a_l a_k > r_k(x,y) dxdy \\ &= \sum_{l=1}^{\infty} \theta_{nl} \sum_{k=1}^{\infty} C_{lk} \int_A r_i(x,y) r_k(x,y) dxdy \\ &= \sum_{l=1}^{\infty} \theta_{nl} \sum_{k=1}^{\infty} C_{lk} R_{ki} \end{aligned}$$

which using the matrix notation becomes

$$\mathbf{A} = \mathbf{C}\boldsymbol{\Theta}^{\mathbf{T}} \tag{4.31}$$

Finally, the last term in Eq (4.27) can be simplified as,

$$S_{np} = < \sum_{l=1}^{\infty} \theta_{nl} a_l \sum_{k=1}^{\infty} \theta_{pk} a_k > + < v_n {v_p}^T >$$

again employing the matrix notation, it becomes,

$$\mathbf{S} = \mathbf{\Theta C \Theta^T + D} \tag{4.32}$$

Thus, the optimal estimator can be written as

$$\mathbf{C\Theta^T [\Theta C \Theta^T + D]^{-1}} \tag{4.33}$$

which can be shown to be equal to the weighted least square equation as follows. Let

$$\mathbf{A} = (\mathbf{\Theta^T D^{-1} \Theta + C^{-1}})^{-1} \tag{4.34}$$

then

$$\begin{aligned} (\mathbf{\Theta^T D^{-1} \Theta + C^{-1}})\mathbf{A} &= \mathbf{I} \\ (\mathbf{\Theta C \Theta^T + D})\mathbf{D^{-1} \Theta A} &= \mathbf{\Theta C} \end{aligned} \tag{4.35}$$

Applying a transpose operator on both sides of Equation (4.35) gives,

$$\begin{aligned} \mathbf{A \Theta^T D^{-1} (\Theta C \Theta + D)} &= \mathbf{C \Theta^T} \\ (\mathbf{\Theta^T D^{-1} \Theta + C^{-1}})^{-1} \mathbf{\Theta^T D^{-1}} &= \mathbf{C \Theta^T (\Theta C \Theta^T + D)^{-1}} \end{aligned} \tag{4.36}$$

Hence the optimal estimator in the technique proposed by Wallner is same as the solution of the weighted least square method.

4.5 MAP Solutions

A final approach to wavefront estimation is the maximum a posteriori technique proposed by Solomon et. al [81]. It is based on the calculation of the maximum a posteriori estimate for the Karhunen-Loeve expansion coefficients. The Bayes theorem states that the posterior probability distribution can be written as a function of the likelihood func-

tion and the a priori law as,

$$\mathbf{P(a|h)} = \frac{\mathbf{P(h|a)P(a)}}{\mathbf{P(h)}} \tag{4.37}$$

and we seek that estimator $\hat{\mathbf{a}}$ which maximises the posterior probability distribution $P(\mathbf{a}|\mathbf{h})$. To do this, both the likelihood function $P(\mathbf{h}|\mathbf{a})$ and the a priori law $P(\mathbf{a})$ have to be found.

The a priori law can be found by first expressing the wavefront as a series of Karhunen-Loeve functions, ie.,

$$\phi(x,y) = \sum_j a_j K_j(x,y) \tag{4.38}$$

where $K_j(x,y)$ is the j^{th} order Karhunen-Loeve function, a_j is the expansion coefficient and $\phi(x,y)$ is the wavefront. According to Noll [7], the expansion coefficients are zero mean Gaussian random variables. Hence, the a priori law of the estimator can be written as,

$$P(\mathbf{a}) = \exp[-\frac{1}{2}\sum_{\mathbf{i}} \frac{\mathbf{a_i}^2}{\mathbf{s_i}}] \tag{4.39}$$

where s_i is the i^{th} eigenvalue of the Karhunen-Loeve covariance matrix. Also by assuming that the point spread function can be approximated by a Gaussian function of mean b_j and variance ${\sigma_j}^2$, the likelihood function can be written as,

$$\begin{aligned} P(\mathbf{h}|\mathbf{a}) &= \prod_j \exp(-\frac{N_j(h_j - b_j)^2}{2{\sigma_j}^2}) \\ &= \prod_j \exp(-\frac{N_j(h_j - \sum_i \theta_{ji}a_i)^2}{2{\sigma_j}^2}) \end{aligned} \tag{4.40}$$

where N_j is the number of detected photons. Using Eq (4.37), Eq (4.39) and Eq (4.40), the posteriori probability can be rewritten as,

$$P(\mathbf{a}|\mathbf{h}) = \exp[-\frac{1}{2}(\sum_{\mathbf{i}} \frac{\mathbf{a_i}^2}{\mathbf{s_i}} + \sum_{\mathbf{j}} \frac{\mathbf{N_j}(\mathbf{h_j} - \sum_{\mathbf{i}} \theta_{\mathbf{ji}}\mathbf{a_i})^2}{{\sigma_\mathbf{j}}^2})] \tag{4.41}$$

By equating the partial derivatives of Eq (4.41) with respect to a particular element of $\mathbf{a}$ to zero and rearranging the results using matrix notation, the solution can be written as,

$$\hat{\mathbf{a}} = [\Theta^T \mathbf{D} \Theta + \mathbf{C^{-1}}]^{-1} \Theta^T \mathbf{D^{-1}} \mathbf{h} \tag{4.42}$$

where $\mathbf{D}$ is,

$$\mathbf{D} = \begin{pmatrix} \frac{N_1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{N_2}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{N_M}{\sigma_M^2} \end{pmatrix} \tag{4.43}$$

and the Karhunen-Loeve covariance matrix $\mathbf{C}$ is,

$$\mathbf{C} = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_N \end{pmatrix} \tag{4.44}$$

The solution thus also reduces to the weighted least square solution, in which the data is weighted through the matrix $\mathbf{D}$ and the a priori knowledge included through the covariance matrix $\mathbf{C}$.

Chapter 5

Practical Limitations

In Chapter 4, a weighted least square approach to the problem of estimating the wavefront distortion is introduced. In this chapter, the practical limitations of correcting the wavefront distortion estimated by the weighted least square approach are discussed. The correction of the distortion by adaptive optics system is limited by several factors:

1. Because the correcting elements, eg., the deformable mirrors, have a finite number of degrees of freedom, the compensation of the wavefront is limited by the ability of the mirror to match the changes to the wavefront introduced by the atmosphere;

2. The wavefront distortion is assumed to be introduced by the Kolmogorov turbulence. In practice, there may be some other contributions to the distortion which are not of Kolmogorov type. Hence the compensation is limited by the accuracy of the Kolmogorov turbulence model;

3. The deformable mirrors are assumed to be Gaussian type as discussed in Chapter 3. However it may not be the case in practice. Thus the compensation is also limited by the accuracy of the modelling of deformable mirrors;

4. the telescope aperture usually is not simply circular but also has central obscuration. This is caused by practical necessity of a secondary mirror in the path of the light received by wavefront sensor. The Shack Hartmann sensor thus cannot provide any measurements from the central part of the telescope aperture, with a consequent degradation of the qualtiy of the estimation.

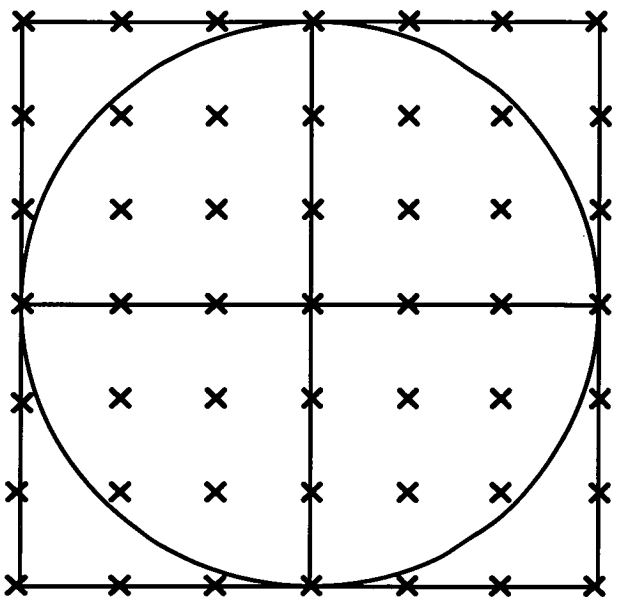


Figure 5.1: The relationship of the actuator array to the sensor array. The actuator position is shown by a 'cross', with the sensor positions shown by 'squares'.

Deviations of the physical system in any of these areas would cause the performance of the wavefront compensation to be degraded. In this section, the effects of these factors are investigated.

5.1 Effect Of Mirrors

Because of the finite number of degrees of freedom in the mirrors, the adaptive optics system may not be able to adjust to produce the required correction to the wavefront. In order to simulate this effect, the adaptive optics system was first modelled as a deformable mirror consisting of a 7×7 array of actuators with a separate tip/tilt mirror (AO1). For comparison, a second adaptive optics system was modelled as a deformable mirror consisting of a 7×7 array of actuators only (AO2). The reason for doing this is that the distortion resulting from tip/tilt is a dominant form of phase distortion compared to other higher order distortions as discussed in Section 4.1. Most practical systems in use have a separate tip/tilt mirror to correct for the tip/tilt distortion and a deformable mirror that corrects for the higher order distortion.

The actuators were assumed to lie on a rectangular grid, with each actuator producing a

Gaussian deflection of the mirror around the actuator centre. The width of the Gaussian produced by the actuator was chosen so that it produced a value of e^{-1} at the adjacent actuator following the suggestion of Wallner [80]. The resulting relationship of the actuator array to the sensor array is shown in Figure 5.1.

The results of simulating 200 phase screens are shown in Table 5.1. The first column of Table 5.1 corresponds to the total number of photons present while the next four columns correspond to the wavefront errors resulting from using different estimation procedures. The second column is the result of using the simple least square method (Eq (4.6)). The third column shows the result of using the weighted least square method presented in Section 4.3 with an assumption of using a perfect mirror. By comparing the second and the third columns, it can be seen that the incorporation of the a priori information of the covariance matrix of the coeffieients of the Zernike polynomials and the noise covariance matrix causes a significant decrease in the residual wavefront error.

In order to investigate the effect of the limited number of degrees of freedom in the mirrors, the optimal results were projected onto the subspace spanned by the two adaptive optics systems. This subspace is defined as the range of all possible wavefronts that can be synthesised using the deformable mirror. The results for these two systems are shown in the fourth and the fifth columns in Table 5.1. It can be seen that the residual errors of either AO1 or AO2 do not deviate much from the optimal estimator, their results are very similar.

In the next two sub-section, the effects of the actuator spacing and the size of lenslet arrays are discussed. It should be noted that the actuator spacing affects the ability of the deformable mirrors to correct for the wavefront distortion while the size of lenslet arrays determines the quanlity of the wavefront estimate from the Shack Hartmann sensor. Simulation results from AO1 and AO2 are also compared.

5.1.1 Effect Of Actuator Spacing

The results in Table 5.1 are obtained for a 7×7 array of actuators in both AO1 and AO2 systems. The similar performance of AO1, AO2 and the optimal estimator is due to the fact that the performance of the adaptive optics system is limited by the sensor measurements, not by the mirrors. As a result, whether having a separate tip/tilt

mirror or not does not affect the resulting residual error. This can be confirmed by increasing the actuator array to 13 × 13. The results are shown in Table 5.2.

However, when the size of the actuator array decreases to 3 × 3 or 5 × 5 , it is apparent from Table 5.3 and Table 5.4 that the performance of AO1 is better than that of AO2. This is because as the adaptive optics system become limited by the ability of the deformable mirror to correct the residual errors. As AO1 has a separate tip/tilt mirror to correct for the residual errors, it has more degrees of freedom than AO2. As a result, AO1 has a better performance than AO2, although not by a large amount

5.1.2 Effect of the size of lenslet

The increase of the size of lenslet array implies that there are more measurements from the Shack Hartmann sensor. In this subsection, the effect of the size of lenslet is investigated. The results of changing the size of lenslet array to 2 × 2, 3 × 3 and 5 × 5 are shown in Tables 5.1, 5.5 and 5.6 respectively. It is apparent that the results of the optimal estimator, AO1 and AO2 systems are very similar despite changing the size of lenslet arrays. Therefore, the quality of reconstruction is not significantly limited by the addition of a tip/tilt mirror. It should be noted that this still may be required to reduce the required displacements in the deformable mirror.

Also, when there are more number of photons available, the residual errors for the optimal reconstructor, AO1 and AO2 decrease despite the size of the lenslet array. It is because when there are more number of photons available, the photon noise is less dominant and thus the estimation is improved.

When comparing the residual errors obtained from different size of lenslet arrays at high photon level (such as 5000), the residual errors for the optimal reconstructor, AO1 and AO2 systems decrease as the size of lenslet array increases. But at low photon level, the residual errors for the optimal estimator, AO1 and AO2 systems generally increases as the size of lenslet array increases. It is because large lenslet array size is useful only when there is sufficient photons onto a sub-aperture. If there is not sufficient photons, the sensor measurements are not reliable, thus causing the residual errors to increase. Hence, there is always a tradeoff between the number of photons onto a sub-aperture and the number of sensor measurements available. If there are lots of photons onto a

		Wavefront error (rad^2)			
		Simple LS	Optimal	AO1	AO2
	25	2.36	2.00	2.00	2.00
	50	1.59	1.42	1.42	1.43
Total	100	1.20	1.07	1.07	1.07
Photons	200	1.01	0.86	0.86	0.86
	500	0.89	0.71	0.72	0.72
	∞	0.81	0.60	0.60	0.61

Table 5.1: The residual wavefront error computed by simulation over 200 phase screens, $D/r0 = 4$ and average variance = 10.17. Measurement errors are assumed to result from photon noise alone. System AO1 has a deformable mirror and a tip/tilt mirror and system AO2 has the deformable mirror only.

Total Photons	25	50	100	200	500	∞
AO1	2.00	1.42	1.07	0.86	0.72	0.60
AO2	2.00	1.42	1.07	0.86	0.72	0.60

Table 5.2: A comparison of the residual errors for AO1 and AO2 by using 13 × 13 actuator arrays.

Total Photons	25	50	100	200	500	∞
AO1	2.00	1.42	1.07	0.86	0.72	0.62
AO2	2.04	1.47	1.12	0.91	0.77	0.67

Table 5.3: A comparison of the residual errors for AO1 and AO2 by using 3 × 3 actuator arrays. Figures quoted are in terms of mean wavefront error in $(rad)^2$. Lenslet array size is 2 × 2.

Total Photons	25	50	100	200	500	∞
AO1	2.00	1.42	1.07	0.86	0.72	0.60
AO2	2.01	1.44	1.08	0.87	0.73	0.62

Table 5.4: A comparison of the residual errors for AO1 and AO2 by using 5×5 actuator arrays.

Total Photons	25	50	100	200	500	1000	5000
optimal	3.02	2.02	1.35	0.92	0.59	0.46	0.33
AO1	3.02	2.02	1.35	0.92	0.59	0.46	0.33
AO2	3.02	2.03	1.35	0.92	0.60	0.46	0.34

Table 5.5: A comparison of the residual errors for the optimal reconstructor, AO1 and AO2 by using 3×3 lenslet arrays and 7×7 actuator arrays.

Total Photons	25	50	100	200	500	1000	5000
optimal	5.32	3.75	2.48	1.59	0.86	0.61	0.24
AO1	5.32	3.75	2.48	1.59	0.86	0.60	0.24
AO2	5.32	3.75	2.49	1.60	0.87	0.60	0.25

Table 5.6: A comparison of the residual errors for the optimal reconstructor, AO1 and AO2 by using 5×5 lenslet arrays and 7×7 actuator arrays.

sub-aperture, one can use a large size of lenslet array to obtain more measurements. Otherwise, a small size of lenslet array should be used so that the sensor measurements are reliable.

5.2 Effect Of $r0$

One of the problems with the Kolmogorov turbulence model is the inaccuracies in estimating the Fried's parameter $r0$. As all the entries of the covariance matrix, $\mathbf{C}$, are proportional to ${r_0}^{5/3}$ (Section 4.3), this implies that $\mathbf{C}$ is affected by the inaccuracies in estimating r_0 and thus the result from the weighted least square estimation is affected. Figure 5.2 and Figure 5.3 show the resulting residual errors when there is an error in estimating r_0 for 2×2 and 4×4 Shack Hartmann sensors under different signal to noise ratios.

It is apparent from Figure 5.2 and Figure 5.3 that at no noise and high signal to noise ratio cases, errors in $\mathbf{C}$ do not have a great impact on the quality of the reconstructed wavefront even when r_0 deviates greatly from its true value. But at low signal to noise ratio, even a small error in estimating r_0 could result in a large increase in the residual error.

5.3 Effect Of Incorrect Estimation Of The Tip/Tilt Coefficients

The other aspects of the deviations from the Kolmogorov turbulence model is the infinite outer scale assumption in the Kolmogorov turbulence. It would cause the covariance of the wavefront tip/tilt coefficient to be over-estimated. As discussed in Section 4.1, the distortion resulting from tip/tilt coefficients is a dominant form of distortion. So if the covariance of the wavefront tip/tilt coefficient is over-estimated, it is most likely that the quality of reconstruction would be degraded significantly. These effects are shown in Figure 5.4 and Figure 5.5. These curves show that the quality of the reconstructed wavefront is significantly affected by the inaccuracies in estimating the tip/tilt coefficient.

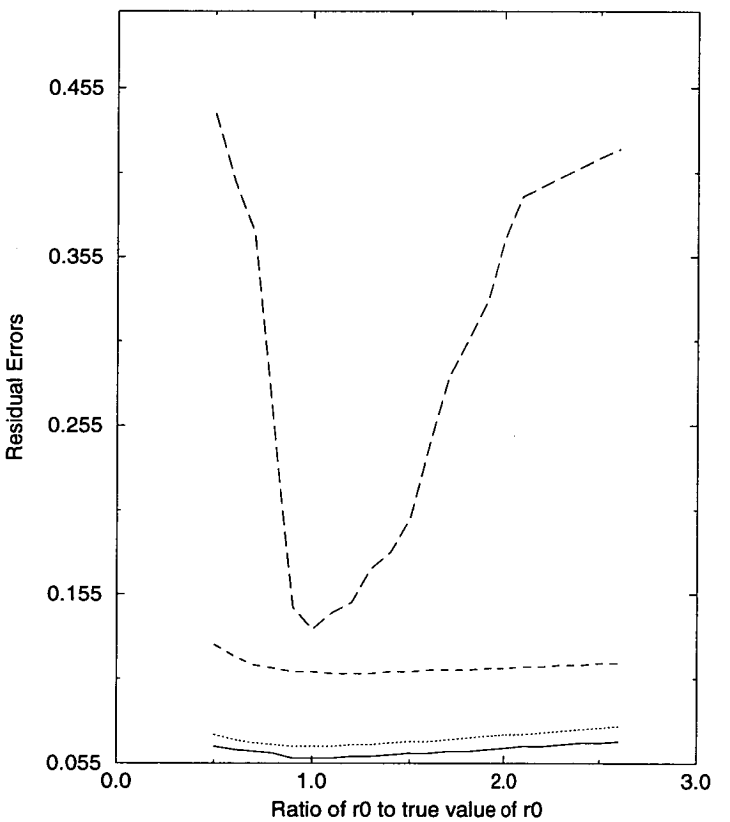


Figure 5.2: The residual errors when there is an error in estimating r_0 for 2×2 Shack Hartmann sensor under different noise level. The solid line, the dotted line, the dashed line and the long dashed line represent no noise, 20 db, 10 db and 0 db SNR cases respectively.

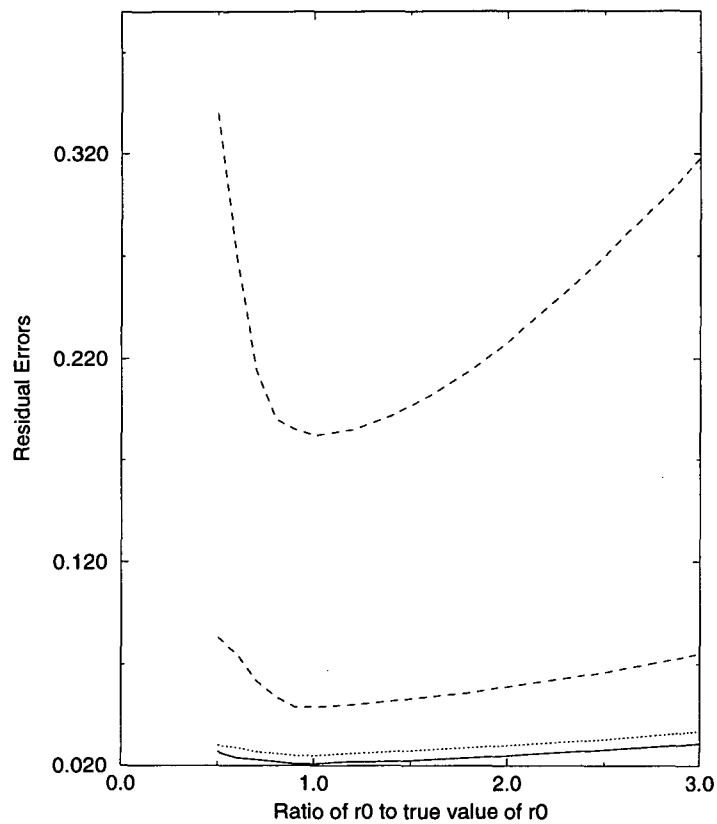


Figure 5.3: The residual errors when there is an error in estimating r_0 for 4×4 Shack Hartmann sensor under different noise level. The solid line, the dotted line, the dashed line and the long dashed line represent no noise, 20 db, 10 db and 0 db SNR cases respectively.

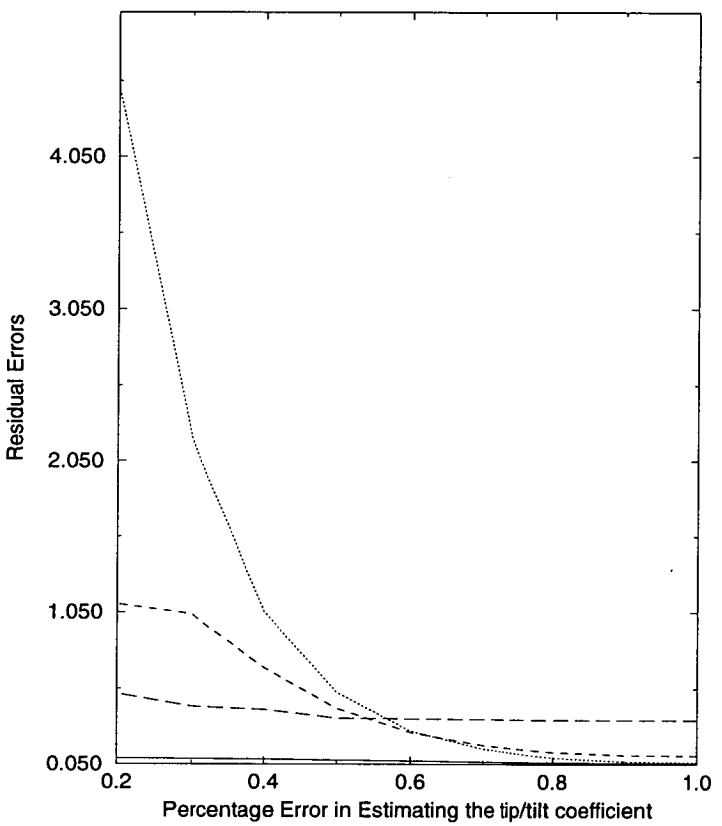


Figure 5.4: The residual errors when there is an error in estimating the tip/tilt coefficient for 2 × 2 Shack Hartmann sensor under different noise level. The solid line, the dotted line, the dashed line and the long dashed line represent no noise, 20 db, 10 db, and 0 db SNR cases respectively.

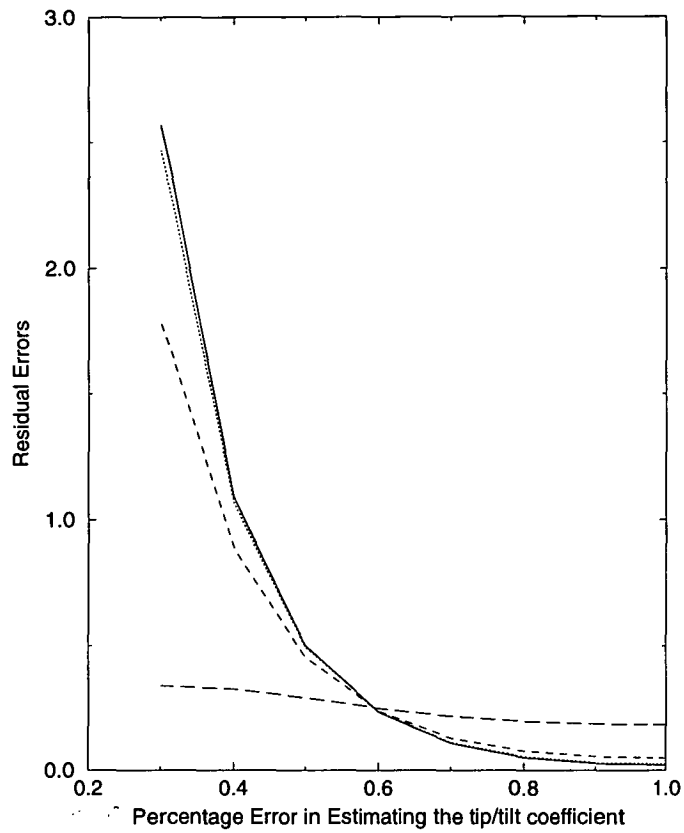


Figure 5.5: The residual errors when there is an error in estimating the tip/tilt coefficient for 4 × 4 Shack Hartmann sensor under different noise level. The solid line, the dotted line, the dashed line and the long dashed line represent no noise, 20 db, 10 db, and 0 db SNR cases respectively.

5.4 Effect Of The Deviation From Kolmogorov Turbulence Model Or a Known Phase Structure Function

The wavefront distortion could also be caused by non-Kolmogorov effects such as the primary mirror not being in thermal equilibrium with the surrounding atmosphere. This is unlikely to produce phase distortion with the structure functions expected of Kolmogorov turbulence model. Fortunately, deviations of either Kolmogorov turbulence or a known phase structure function can be replaced by the assumption that the a priori covariance of the actuator functions is known. Assuming the statistics of the turbulence remain constant for a short period of time, this data can be obtained by observing a nearby bright star. The nearby bright star have sufficient photons so that a large lenslet array could be employed without noise limitation in order to estimate the covariance of the actuator functions for use in the algorithm.

In our study, we employed a 8×8 lenslet Shack-Hartmann array which produced for $D/r0 = 4$ a residual error variance of $6.03 \times 10^{(-2)}$. This was then projected onto the deformable mirror/fixed mirror subspace (AO1 system) resulting in a residual error of $8.3 \times 10^{(-2)}$. A simulation of 1000 observations was then performed and the covariance of the actuator functions was computed. This covariance along with the interaction matrix computed for the actual mirror functions were used to reconstruct the wavefront. The simulation results are shown in Table 5.4. As can be seen from Table 5.4, the performance of using the covariance of the actuator functions does not differ significantly from using Wallner's method (Chapter 4) which require the assumption of a structure function.

5.5 Section Conclusion

In this section, various factors affecting the quality of reconstruction were discussed. It is found that the estimator is insensitive to the errors in r_0 at high signal to noise ratio case but is sensitive to the errors in r_0 at low signal to noise ratio case. Also, it is found that the inaccuracies in estimating the tip/tilt coefficients can have a significant impact on the quality of the reconstruction.

A tradeoff between the size of the lenslet array and the quality of the estimation is

		Residual errors	
		optimal	Using the observed covariance and AO1
	25	2.00	1.99
	50	1.42	1.42
Total	100	1.07	1.06
Photons	200	0.86	0.85
	500	0.71	0.71
	∞	0.60	0.59

Table 5.7: A comparison of the residual wavefront errors between the optimal estimator and the one using the observed covariance and AO1 system.

noted. If there are lots of photons onto the aperture, then a large size of lenslet arrays could be employed. Otherwise, a small size of the lenslet arrays should be used due to the noise limitation. It is also shown that the required a priori information can be obtained by observation with adaptive optics system. In particular, by observing a bright object with a large lenslet array, the covariance of the actuator functions can be calculated. The results attained could then be employed to advantage as the a priori information about the covariance of the actuator functions for fainter objects where the size of the lenslet array is restricted by the available photons.

Chapter 6

Blind Deconvolution

The recovery of both the object, $f(x,y)$, and the point spread function, $h(x,y)$, from the blurred image, $g(x,y)$, poses a blind deconvolution problem. This problem is significantly more demanding and difficult to solve than the conventional deconvolution problem in which $h(x,y)$ is known. The difficulty of solving this problem is, however, reduced by applying some a priori knowledge about the general properties of $f(x,y)$ and $h(x,y)$. For a general imaging problem, these a priori knowledge could be:

1. the support constraint that restricts the size of an image to a defined region;
2. the positivity constraint that restricts the image intensity to be positive.

In some cases, it is possible to obtain several statistically independent blurred images of the same object. For example, in the case of astronomical imaging, a series of short-exposure images can be recorded at ground level. Due to the time-varying nature of the atmospheric turbulence, each short-exposure image has the common object but is blurred by a different point spread function (Section 2.3). This is an important a priori constraint because the estimation process constrains all the image frames to have a common object but permits the point spread function and noise to vary with each image. The multiple blurring of the same object helps to reduce the inherent ambiguities in the blind deconvolution problem as discussed in Section 2.4.

The basic model of deconvolution we consider has already been given in Eq (2.13), and is repeated here as,

$$g_n(x,y) = f(x,y) \odot h_n(x,y) + c_n(x,y) \tag{6.1}$$

or alternatively, in Fourier domain as,

$$G_n(u,v) = F(u,v)H_n(u,v) + C_n(u,v) \tag{6.2}$$

The blind deconvolution problem can easily be formulated using the Bayes theorem which states that [88],

$$Pr(h \odot f|g) = \frac{Pr(g|h \odot f)Pr(f)Pr(h)}{Pr(g)} \tag{6.3}$$

Maximizing the term $Pr(h \odot f|g)$ is equivalent to minimizing ϵ which is defined as,

$$\epsilon = -\log[Pr(g|h \odot f)] - \log[Pr(f)] - \log[Pr(h)] \tag{6.4}$$

In Eq (6.4), the term $\log[Pr(g|h \odot f)]$ represents the convolution requirement, ie., the convolution of $f(x,y)$ and $h(x,y)$ to be consistent with the blurred image $g(x,y)$ while the other two terms, $\log[Pr(f)]$ and $\log[Pr(h)]$, represent the a priori knowledge about the object and the point spread function respectively.

In order to evaluate an expression for $\log[Pr(g|h \odot f)]$, the noise statistics need to be known. If the noise is assumed to follow a Poisson distribution, then $\log[Pr(g|h \odot f)]$ can be written as,

$$\log[Pr(g|h \odot f)] = -\sum_{x,y} h(x,y) \odot f(x,y) + \sum_{x,y} g(x,y) \log[h(x,y) \odot f(x,y)] \tag{6.5}$$

If the noise is assumed to follow a Gaussian distribution, then it is written as,

$$\log[Pr(g|h \odot f)] = -\frac{1}{2}\sum_{x,y}[\frac{h(x,y) \odot f(x,y) - g(x,y)}{\sigma(x,y)}]^2 \tag{6.6}$$

where $\sigma(x,y)$ is the standard deviation of $g(x,y)$.

There are two philosophies for incorporating the a priori information about the object and the point spread function into the restoration process. The first is to assume that the a priori knowledge is absolute. This means that the estimators, $f(x,y)$ and $h(x,y)$,

must satisfy the knowledge exactly. The other philosophy to incorporate the knowledge is to regard the knowledge as inexact and we try to find a solution that would satisfy all the a priori knowledge in the form of an error if possible. If the a priori knowledge have some conflicting requirements, then the algorithm tries to find a solution that would minimize the error resulting from not satisfying this knowledge.

In this section, three different blind deconvolution algorithms are discussed. These are the conjugate gradient minimization approach, the maximum likelihood approach and the projection based approach. All these three methods are based on Eq (6.4) but with different assumption made on the distribution of the noise statistics and a different application of the a priori knowledge about $f(x,y)$ and $h(x,y)$.

6.1 Conjugate Gradient Minimization Approach

As stated in Section 3.4, one of the disadvantages of using the iterative blind deconvolution loop is that it cannot be equated to a steepest descent search on an error metric in either the Fourier or image space. As a result, performing an extra iteration of the algorithm may result in new estimates of $f(x,y)$ and $h(x,y)$ which may be worse than the estimates before the iteration. In order to solve this convergence problem, the estimation problem can be redefined as an unconstrained minimization problem [78,89]. Minimizing an error metric is a common technique to be applied in many different situations.

There are two types of knowledge. The first is our observed data $g(x,y)$. The second is a statistical knowledge of the form of $f(x,y)$ and $h(x,y)$. These can be formulated as two error terms, namely the convolution error and the image error. The latter is the estimated values of $f(x,y)$ and $h(x,y)$ that violate their respective a priori information in image space, and the former is the square difference between the product of the estimated values of $f(x,y) \odot h(x,y)$ and the given measurement $g(x,y)$. An overall error to be minimized can then be formed by combining the image space error and the convolution error, ie.,

$$E_t = E_c + \beta E_i \tag{6.7}$$

where

$$E_i = \sum_{\gamma_f} |f(x,y)|^2 + \sum_{j=1}^{N} \sum_{\gamma_{h_j}} |h_j(x,y)|^2 \tag{6.8}$$

and

$$\begin{aligned} E_f &= \sum_{j=1}^{N} \sum_{x} \sum_{y} (g_j(x,y) - f(x,y) \odot h_j(x,y))^2 \\ &= \sum_{j=1}^{N} \sum_{u} \sum_{v} |G_j(u,v) - F(u,v)H_j(u,v)|^2 \end{aligned} \tag{6.9}$$

γ_f and γ_{h_j} are pixels where $f(x,y)$ and $h_j(x,y)$ violates their respective image space constraints, N is the total number of multiple image frames available and β is the weighting between the convolution error and the a priori error. E_c represents the convolution error, ie., the derivation of the estimates $F(u,v)$ and $H_j(u,v)$ from the known constraint $G_j(u,v)$ in Fourier space. E_i represents the a priori error of the estimates $f(x,y)$ and $h_j(x,y)$. So by putting all the constraints into the overall error, one can minimize it to obtain good estimates of both the object and the point spread function.

When comparing Eq (6.9) and Eq (6.6), it can be seen that the least square formulation of the convolution error is actually equivalent to the Bayes formulation with an assumption of Gaussian noise and the standard deviation of $g_j(x,y)$ to be one.

A conjugate gradient minimization routine is chosen to minimize E_t. This algorithm is a robust minimization routine with well-defined termination conditions. In practice, the routine produces a series of estimates of $f_k(x,y)$ and $h_{j_k}(x,y)$ having a monotonically decreasing value of E_t with increasing number of iterations k [90]. Convergence occurs when the algorithm has found a global or a local minimum in this function, i.e., when any perturbation of the estimated pixels of either $f(x,y)$ or $h_j(x,y)$ causes an increase in E_t.

In order to apply the conjugate gradient routine effectively, we have to calculate the rate of change of E_t with respect to each of the variables, ie., pixels which comprise $f(x,y)$ and $h_j(x,y)$. The derivatives of E_i with respect to these pixels can be written

as :

$$\frac{\partial E_i}{\partial f(n,m)} = \begin{cases} 2f(n,m) & \text{if } f(n,m) \text{ violates some a priori information} \\ 0 & \text{otherwise} \end{cases} \tag{6.10}$$

and

$$\frac{\partial E_i}{\partial h_j(n,m)} = \begin{cases} 2h_j(n,m) & \text{if } h_j(n,m) \text{ violates some a priori information about } h_j(n,m) \\ 0 & \text{otherwise} \end{cases} \tag{6.11}$$

The derivatives of E_f with respect to $f(n,m)$ and $h_j(n,m)$ can be written as,

$$\begin{aligned} \frac{\partial E_f}{\partial f(n,m)} &= -2\sum_{j=1}^{N}\sum_{x,y} h_i(n-x,m-y) \\ &\quad (g_j(x,y) - \sum_{x',y'} h_j(x'-x,y'-y)f(x',y')) \end{aligned} \tag{6.12}$$

and

$$\begin{aligned} \frac{\partial E_f}{\partial h_j(n,m)} &= -2\sum_{x,y} f(n-x,y-m) \\ &\quad (g_j(x,y) - \sum_{x',y'} f(x'-x,y'-y)h_j(x',y')) \end{aligned} \tag{6.13}$$

As a problem stated by Yang, Galatsanos and Stark [91], the overall error that is minimized by the conjugate gradient routine yields different values for the equivalent pair of solutions $(f, h_1, h_2, \cdots, h_N)$ and $(\alpha f, \frac{h_1}{\alpha}, \frac{h_2}{\alpha}, \cdots, \frac{h_N}{\alpha})$, where α is a positive constant. The conjugate gradient approach described above is thus sensitive to the brightness level. As can be seen from Eq (6.7), the overall error consists of two parts. The first part, E_c, is unaffected by the scale constant α, but the second part, E_i, is affected. For the solution $(\alpha f, \frac{h_1}{\alpha}, \frac{h_2}{\alpha}, \cdots, \frac{h_N}{\alpha})$, the image space error becomes,

$$E_i = \alpha^2 \sum_{\gamma_f} |f(x,y)|^2 + \frac{1}{\alpha^2}\sum_{j=1}^{N}\sum_{\gamma_{h_j}} |h_j(x,y)|^2 \tag{6.14}$$

Hence α changes the weightings between the image space error of the estimated object and that of the estimated point spread function. If α is large, the image space error in the estimated object is dominant which implies that the error resulting from the estimated point spread function is less important than that from the estimated object.

Thus the estimations obtained is not sensitive to the error in the point spread function even though $h_k(x, y)$ violates its constraints.

This scale problem can be solved by normalizing the initial estimates of both $f(x, y)$ and $h(x, y)$ to be of equal energy within their respective support constraints [79,89]. Also, it should be noted that the relative importance of E_c and E_i is not fixed in this algorithm, a feature which can be used to advantage when dealing with the inherently variable nature of actual data. This results from the spectrum of the estimated convolution being the product of $\hat{F}(u, v)$ and $\hat{H}(u, v)$. If we scale the observed convolution by ζ, the components only increase by $\sqrt{\zeta}$. This changes the function being minimised to

$$E_t^{new} = \zeta^2 E_c + \zeta E_i, \tag{6.15}$$

which results in a greater emphasis on the convolutional error with respect to support and positivity errors. Conversely in order to enforce the support and positivity constraints more exactly, we need merely scale the convolution down.

In summary, the conjugate gradient approach has a flexibility in specifying the relative importance of the convolution error and the image space error. Also, in this approach, the a priori information, ie., the image space and the Fourier space constraints, are not enforced absolutely, which implies that the algorithm tries to find a solution that would minimize the error resulting from violating these a priori information.

6.2 Maximum Likelihood Approach

Another technique to solve the blind deconvolution problem is the maximum likelihood approach proposed by Schulz [92]. It is a probabilistic approach to solve the problem. It tries to find the object and the point spread function that are most likely to have created the observed blurred image.

In order to define a likelihood function, the statistical nature of the noise needs to be known. As the mechanism of photon arrival in the image plane is always modelled as a Poisson process [5], the dominant form of noise considered here is photon noise and $g_j(x, y)$ is assumed to be a Poisson-distributed random variable with mean equal to $f(x, y) \odot h_j(x, y)$. The log likelihood function can thus be defined as shown in Eq (6.5)

for single frame case. As for multiple frame case, it is written as,

$$L(f,h) = \sum_{j=1}^{N}[-\sum_{x,y} f(x,y) \odot h_j(x,y) + \sum_{x,y} g_j(x,y) \log[f(x,y) \odot h_j(x,y)]] \quad (6.16)$$

The blind deconvolution problem is then solved by the technique of maximum likelihood which has been discussed in Section 1.4. The problem can be stated as finding $\hat{f}$ and $\hat{h}$ such that

$$L(\hat{f},\hat{h}) \geq L(f,h) \quad (6.17)$$

for all admissible choices of f and h.

In the presence of noise, maximizing the log likelihood function defined in Eq (6.16) would almost certainly give a trivial solution (Section 2.4). If one chooses $\hat{f}(\mathbf{x})$ to be a Dirac delta function and $\hat{h}(\mathbf{y}|\mathbf{x})$ to be $g_k(\mathbf{y})$, then it is easy to show that this choice maximizes the log likelihood function. This trivial estimate, however, does not contain any information about the object unless the object is truly a point source.

In order to avoid this trivial solution, the a priori information that the object itself is not a delta function needs to be incorporated into the maximum likelihood estimation. Schulz [92] thus puts an extra term into the log likelihood function which he called it a penalty function. This penalty function is used to remove the possibility of estimating the object as a point source. The penalized log likelihood function is then defined as,

$$L_\phi(f,h) = L(f,h) - \beta\phi(f) \quad (6.18)$$

where β is a parameter that specifies how strong the penalty is applied and the penalty function is chosen so that the estimated object is not a delta function and is defined as follows,

$$\phi(f) = -\sum_{x,y} \log[1 - f(x,y)] \quad (6.19)$$

Thus when $f(x,y)$ is a delta function, the penalty function is large and the penalized log likelihood function is certainly not maximized by this choice. But for other choices of $f(x,y)$, the penalty function is small. It should be noted that this choice of the penalty function is quite arbitrary.

It should be noted that in order for Eq (6.19) to be evaluated numerically, $f(\mathbf{x})$ has to

be less than 1. But as $f(x,y)$ is non-negative, so the feasible range of $f(x,y)$ is between 0 and 1. Also it should be noted that the log likelihood function is unaffected by the brightness level, ie., both $(f, h_1, h_2, \cdots, h_N)$ and $(\alpha f, \frac{h_1}{\alpha}, \frac{h_2}{\alpha}, \cdots, \frac{h_N}{\alpha})$ have the same log likelihood value where α is a positive scalar. Hence one can restrict the estimate f to be a member of the set $\mathbf{A}$, which is define as,

$$\mathbf{A} = \{\mathbf{f} : \sum_{\mathbf{x},\mathbf{y}} \mathbf{f(x,y)} = \mathbf{1}, \mathbf{f} \geq \mathbf{0}\} \tag{6.20}$$

A direct and closed form solution is always difficult. Schulz uses an iterative solution based on the GEM algorithm (Section 1.4). The algorithm provides a sequence of estimates of the object and the point spread function that are updated in a manner such that the log likelihood value is non-decreasing, ie.,

$$L_\phi(f^{new}, h^{new}) \geq L_\phi(f^{old}, h^{old}) \tag{6.21}$$

where the superscripts *old* and *new* denote the current and the next estimate. The next estimates of the object and the psf are found to be,

$$h_j{}^{new}(x,y) = h_j{}^{old}(x,y) \sum_{x',y'} \frac{f^{old}(y'-y, x'-x) g_j(x',y')}{f^{old}(x',y') \odot h_j{}^{old}(x',y')} \tag{6.22}$$

$$f^{new}(x,y) = \begin{cases} \frac{\bar{f}(x,y)}{\sum_{x,y} \bar{f}(x,y)} & \text{if } Q_f(\frac{\bar{f}(x,y)}{\sum_{x,y} \bar{f}(x,y)} | f^{old}, h^{old}) \geq Q_f(f^{old} | f^{old}, h^{old}) \\ f^{old}(\mathbf{x}) & \text{otherwise} \end{cases} \tag{6.23}$$

where

$$Q_f(f | f^{old}, h^{old}) = \sum_{x,y} \mu(x,y) \log(f(x,y)) + \beta \sum_{x,y} \log(1 - f(x,y)) \tag{6.24}$$

$$\mu(x,y) = \sum_{j=1}^{N} \sum_{x',y'} \frac{h_j{}^{old}(x'-x, y^prime - y) f^{old}(x,y)}{f^{old}(x',y') \odot h_j{}^{old}(x',y')} g_j(x',y') \tag{6.25}$$

and

$$\bar{f}(x,y) = \frac{\mu(x,y) + D + \beta - \sqrt{[\mu(x,y) + D + \beta]^2 - 4D\mu(x,y)}}{2D} \tag{6.26}$$

By updating the estimated object and the estimated point spread function according to the above equations, the penalized log likelihood value increases with the increasing number of iterations. It should be noted that the a priori knowledge about the support

and the positivity information are applied absolutely in this algorithm. If the algorithm is started with some of the pixels zero, ie., those outside the support region, these pixels remain to be identically zero throughout all the iterations. Also, when f^{old} is greater than zero, the new estimate f^{new} is also greater than zero by Eq (6.23). Thus the positivity constraint is automatically enforced absolutely. There is, however, no guarantee that the algorithm will converge to a global maximum of the likelihood function.

6.3 Projection Based Approach

Another technique to solve the blind deconvolution problem is based on the idea of alternating projection between some constraint sets which are formed by considering some a priori information about the object and the point spread function [22,27]. Two constraint sets based on the a priori knowledge about $f(x,y)$ and $h(x,y)$, denoted by C_f and C_h respectively are constructed. C_f is a set containing all points of $f(x,y)$ that satisfy its a priori information, ie., it contains all points except those in γ_f and C_h is a set containing all points except those in γ_h (Section 1.4).

In this approach, the blind deconvolution is reformulated as a standard minimization problem but subjected to two constraint sets C_f and C_h. It is similar to the conjugate gradient approach since both are basically a minimization problem but in this algorithm the a priori information is enforced differently. In this approach the a priori information is enforced absolutely by the projection operator while the information in the conjugate gradient approach is applied loosely by minimizing the image space error.

The quantity that needs to be minimized is the convolution error in the conjugate gradient approach and is repeated here as,

$$J_{(f,h_1,\ldots,h_N)} = \sum_{j=1}^{N} (g_j(x,y) - f(x,y) \odot h_j(x,y))^2 \tag{6.27}$$

The algorithm is outlined as below.

1. Take an initial guess of $(f, h_1, h_2, \ldots, h_N)$.

2. For k^{th} iteration and the j^{th} frame image, choose $\{h_j\}_k \in C_{h_j}$ so that

$$J_{(f_{k-1},h_{(1,k)},\ldots,h_{(j,k)},\ldots,h_{(N,k-1)})} \leq J_{(f_{k-1},h_{(1,k)},\ldots,h_{(j,k-1)},\ldots,h_{(N,k-1)})} \tag{6.28}$$

   Repeat Step 2 for j from 1 to N.

3. Choose $f_k \in C_f$ so that

$$J_{(f_k,h_{(1,k)},\ldots,h_{(N,k)})} \leq J_{(f_{k-1},h_{(1,k)},\ldots,h_{(N,k)})} \tag{6.29}$$

4. Set $k = k + 1$, repeat Step 2 and Step 3 until convergence is achieved.

Note that, according to the above algorithm, the cost function J satisfies the following inequality,

$$J_{(f_k,h_{1,k},\ldots,h_{N,k})} \leq J_{(f_{k-1},h_{1,k},\ldots,h_{N,k})} \leq \cdots \leq J_{(f_{k-1},h_{1,k},\ldots,h_{N,k-1})} \leq J_{(f_{k-1},h_{1,k-1},\ldots,h_{N,k-1})} \tag{6.30}$$

Hence, by repeated application of this algorithm, the non-increasing sequence of the value of J converges either to zero or to a local minimum [93]. In Step 2, the following iteration will yield a minimum: for the j^{th} frame image in k^{th} iteration,

1. set $h^0 = h_{j,k-1}$;

2. do M times (l = 0 to M-1)

$$h^{l+1} = P_h[h^l - \alpha_k \bigtriangledown J_{(f_{k-1},h_{(1,k)},\ldots,h^l,\ldots,h_{(N,k-1)})}] \tag{6.31}$$

   where $\bigtriangledown$ is the gradient operator, and α_k is a constant chosen to guarantee convergence;

3. set $h_{(j,k)} = h^M$.

In order to determine the value of α_k, Eq (6.27) is rewritten in vector-matrix notation as,

$$J = \sum_{j=1}^{N} ||\mathbf{g_j} - \mathbf{F_{k-1}}\mathbf{h_j}||^2 \tag{6.32}$$

where $\mathbf{g_i}$ and $\mathbf{h_j}$ are vectors. $\mathbf{F_{k-1}}$ is a circulant/Toeplitz matrix of the shifted $\mathbf{f_{k-1}}$. The gradient of J with respect to $\mathbf{h_i}$ can be computed to be,

$$\nabla J = 2\mathbf{F}_{\mathbf{k-1}}^{\mathbf{T}}\mathbf{F_{k-1}}\mathbf{h_i} - 2\mathbf{F}_{\mathbf{k-1}}^{\mathbf{T}}\mathbf{g_i} \tag{6.33}$$

where a superscript, $\mathbf{T}$, denotes matrix transpose. To guarantee the convergence of Eq (6.31) , the constant α_k has to be chosen such that the non-linear mapping given by

$$G(h) = h - \alpha_k \nabla J \tag{6.34}$$

is a contraction mapping, ie., there exists θ where $0 < \theta < 1$ and $h_1 \neq h_2$ such that

$$||G(h_1) - G(h_2)|| \leq \theta ||h_1 - h_2|| \tag{6.35}$$

Due to the fact that $\mathbf{F^T F}$ is a symmetrical matrix, it can be shown that α_k is between 0 and $\frac{1}{\lambda_k}$ where λ_k is the spectral radius of $\mathbf{F}_{\mathbf{k-1}}^{\mathbf{T}}\mathbf{F_{k-1}}$.

In Step 3, a similar iteration is adopted as follows:

1. set $f^0 = f_{k-1}$,
2. do M times (l = 0 to M-1)

$$f^{l+1} = P_f[f^l - \beta_k \nabla J_{(f^l, h_{(1,k)}, \ldots, h_{(N,k)})}] \tag{6.36}$$

3. set $f_k = f^M$

Eq (6.27) can also be rewritten as,

$$J = \sum_{j=1}^{N} ||\mathbf{g_j} - \mathbf{H_j f}||^2 \tag{6.37}$$

where $\mathbf{H_j}$ is a circulant/Toeplitz matrix of the shift $\mathbf{h_i}$. The gradient of J with respect to $\mathbf{f}$ can be shown to be,

$$\nabla J = 2\sum_{j=1}^{N} \mathbf{H}_{\mathbf{j}}^{\mathbf{T}}\mathbf{H_j f} - 2\sum_{\mathbf{j=1}}^{\mathbf{N}} \mathbf{H}_{\mathbf{j}}^{\mathbf{T}}\mathbf{g_j} \tag{6.38}$$

Similarly, in order to ensure contraction mapping, it can be proved that β_k should be

between 0 and $\frac{1}{\lambda_k}$ where λ_k is the spectral radius of $\sum_{j=1}^{N} \mathbf{H_j^T H_j}$.

6.3.1 Addition of A Momentum Term

Note that both Step 2 and Step 3 are basically a steepest descent search. As noted by many researchers [63,94], the steepest descent search is slow to converge (Section 1.4). Hence, this algorithm can take many iterations to converge to the final solution. Here the use of a momentum term in the minimization so as to improve the convergence rate is proposed. The idea of using a momentum term originated from the field of neural network where the momentum term is used to speed up the convergence rate [95–98].

Momentum implements the heuristics through the addition of a new term to the image update equations [99]. For example in Step 2, h is updated by the following equation,

$$h^{l+1} = P_h[h^l + \triangle h_l] \tag{6.39}$$

where

$$\triangle h_l = -(1-\delta)\alpha_k \bigtriangledown J_{(f_{k-1},h^l)} + \delta(h^l - h^{l-1}) \tag{6.40}$$

and δ is the momentum factor that determines the relative contribution of the current partial derivative of J with respect to h and the past changes in h. The search reduces to the steepest descent search when δ equals to zero. The momentum term is used to build in some inertia, ie., the current change should be somewhat similar to the change undertaken at the previous step. By ignoring the projection operator P_h, Eq (6.40) can be rewritten as,

$$\triangle h_l = -(1-\delta)\alpha_k \sum_{i=1}^{l} \delta^{l-i} \bigtriangledown J_{(f_{k-1},h^i)} + \delta^l \triangle h_0 \tag{6.41}$$

The use of the momentum term is justified by the fact that when consecutive partial derivatives of J with respect to h possess the same sign, the RHS of Eq (6.41) gets larger in magnitude and hence h is adjusted by a relatively large amount. Similarly, when consecutive derivatives of J with respect to h possess opposite signs, the summation in Eq (6.41) becomes small in magnitude and h is adjusted by a small amount. By choosing an appropriate choice of δ, the convergence rate can be improved with the use of a momentum term in both Step 2 and Step 3.

6.3.2 Adaptive Step Size Algorithm

The adaptive step size algorithm updates both the images, ie., f and h and the step size, ie., α_k and β_k. In Step 2, the image update equation can be written as,

$$h^{l+1} = P_h[h^l - \alpha^l \bigtriangledown J_{(f_{k-1},h^l)}] \tag{6.42}$$

where α^l is defined as[98],

$$\alpha^l = \alpha^{l-1} + \begin{cases} c & \text{if } \bar{\gamma}^{l-1} \bigtriangledown J_{(f_{k-1},h^l)} > 0 \\ -\phi\alpha^{l-1} & \text{if } \bar{\gamma}^{l-1} \bigtriangledown J_{(f_{k-1},h^l)} < 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.43}$$

and

$$\begin{aligned} \bar{\gamma}^l &= (1-\theta) \bigtriangledown J_{(f_{k-1},h^l)} + \theta\bar{\gamma}^{l-1} \\ &= (1-\theta) \sum_{i=0}^{l} \theta^{l-i} \bigtriangledown J_{f_{k-1},h^i)} \end{aligned} \tag{6.44}$$

Note that $\bar{\gamma}^l$ is a weighted sum of the current and past partial derivatives of J with respect to h according to Eq (6.44). θ is the weighting between two consecutive partial derivatives of J. Hence according to this algorithm, if the current derivative of J with respect to h and the weighted sum of all the previous derivatives of J with respect to h possess the same sign, then α is incremented by a positive constant c so that h can be adjusted by a relatively large amount. If the current derivative of J with respect to h and the weighted sum of all the previous derivatives of J with respect to h possess opposite signs, then α is decremented by a proportion, ϕ, of its current value. Hence, h is adjusted by a small amount. Similarly, f and β_k in Step 3 are updated using the same procedure described above. By employing this strategy, again the convergence rate can be improved.

Chapter 7

Comparative Evaluation

Three blind deconvolution algorithms are discussed in Chapter 6. These three algorithms can be formulated using the Bayes theorem but with different assumption on the noise statistics and different application of the a priori knowledge about the object and the point spread function. The comparative success of these methods for restoring images, measured in the quality of the reconstruction and the computational cost to obtain the reconstruction, is important when one is looking for some methods to solve any particular blind deconvolution problem. In this chapter, the performance of these blind deconvolution algorithms is evaluated. First of all, we show, by using some simulation results, how the use of the momentum term and the use of the adaptive step size improve the convergence rate in Section 7.1. Then, the results of maximum likelihood and the conjugate gradient approaches are examined in Section 7.2 and Section 7.3 respectively. Finally, the three techniques for blind deconvolution problem are compared.

As pointed out in Section 2.4.2, there may exist a number of mathematical valid solutions to a blind deconvolution problem. In this case, it is extremely unreliable to rely on simply one starting point. Only when the same solution is reached from different starting points can one have confidence in the solution achieved. Hence, in the following simulation, five differently initialized initial estimates are used to test the algorithm.

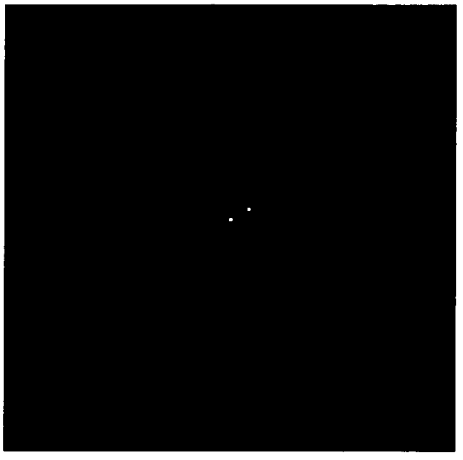


Figure 7.1: The original object used in the simulation. It is a two-point source object with 64×64 *dimension.*

7.1 Accelerated Projection Based Algorithm

As mentioned in Section 6.3, the original projection-based algorithm is a steepest descent approach to the blind deconvolution problem. There are two methods to improve the convergence rate of the projection-based algorithm, namely the use of the momentum term and the use of the adaptive step size. In this section, the performance of these two methods compared with the original projection-based algorithm is investigated. The original object used is a two-point source object and is shown in Fig 7.1. It is degraded by a fractal-like time varying turbulence. Fig 7.2 shows a set of the degraded images. It should be noted that the four degraded images are assumed to be statistically independent as discussed in Chapter 2. We set M to 50 and perform a total number of 500 iterations.

7.1.1 Use of the Momentum Term

The momentum term is used to build in inertia to the algorithm, ie., the current change is assumed to be similar to the change undertaken at the previous step. This term thus controls the weighting between the current gradient and the previous change. If it is zero, then the search reduces to the steepest descent search and the previous change is ignored totally. It it is one, then the current change depends completely on the previous change and the current gradient is ignored.

Figure 7.1: The original object used in the simulation. It is a two-point source object with 64 × 64 dimension.

7.1 Accelerated Projection Based Algorithm

As mentioned in Section 6.3, the original projection-based algorithm is a steepest descent approach to the blind deconvolution problem. There are two methods to improve the convergence rate of the projection-based algorithm, namely the use of the momentum term and the use of the adaptive step size. In this section, the performance of these two methods compared with the original projection-based algorithm is investigated. The original object used is a two-point source object and is shown in Fig 7.1. It is degraded by a fractal-like time varying turbulence. Fig 7.2 shows a set of the degraded images. It should be noted that the four degraded images are assumed to be statistically independent as discussed in Chapter 2. We set M to 50 and perform a total number of 500 iterations.

7.1.1 Use of the Momentum Term

The momentum term is used to build in inertia to the algorithm, ie., the current change is assumed to be similar to the change undertaken at the previous step. This term thus controls the weighting between the current gradient and the previous change. If it is zero, then the search reduces to the steepest descent search and the previous change is ignored totally. It it is one, then the current change depends completely on the previous change and the current gradient is ignored.

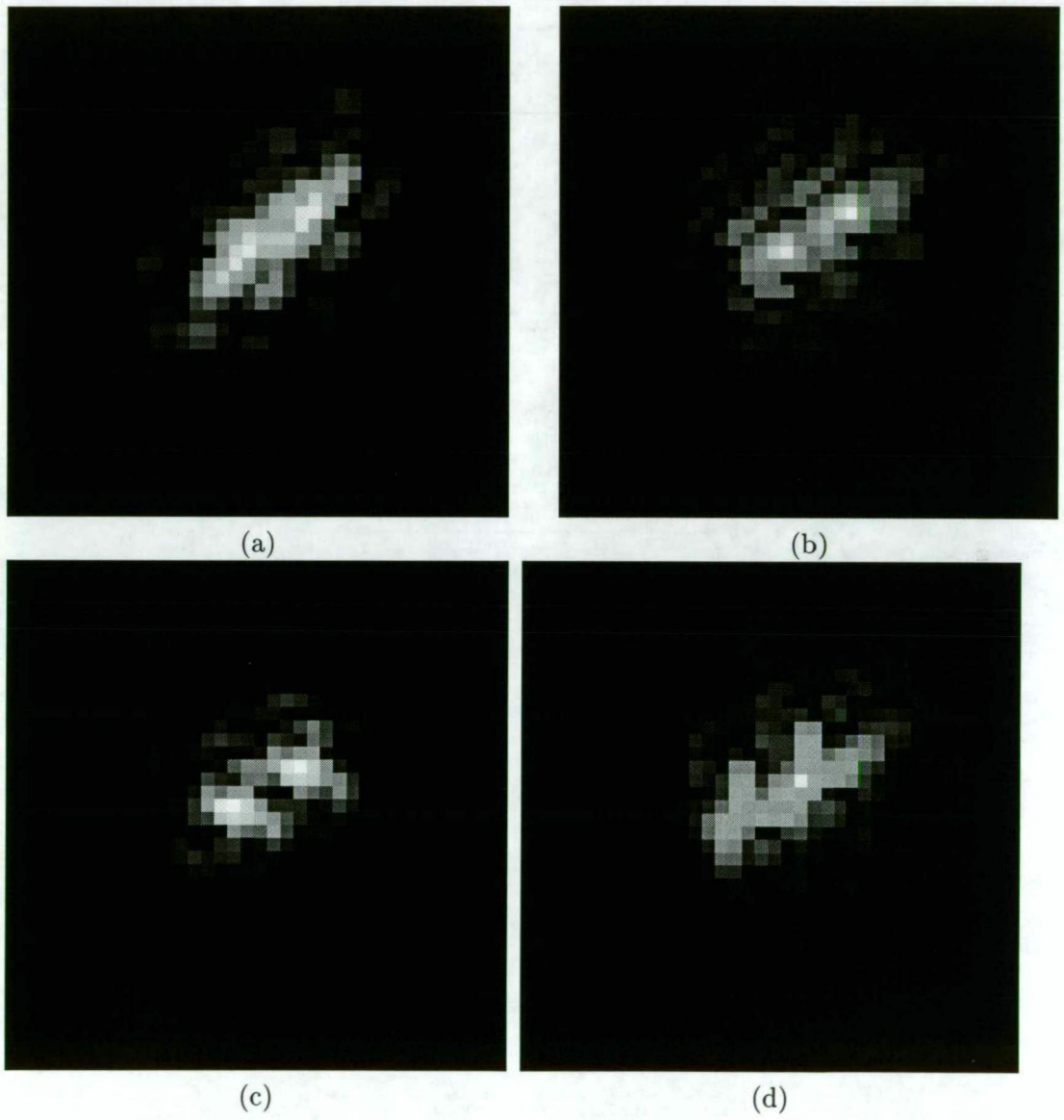
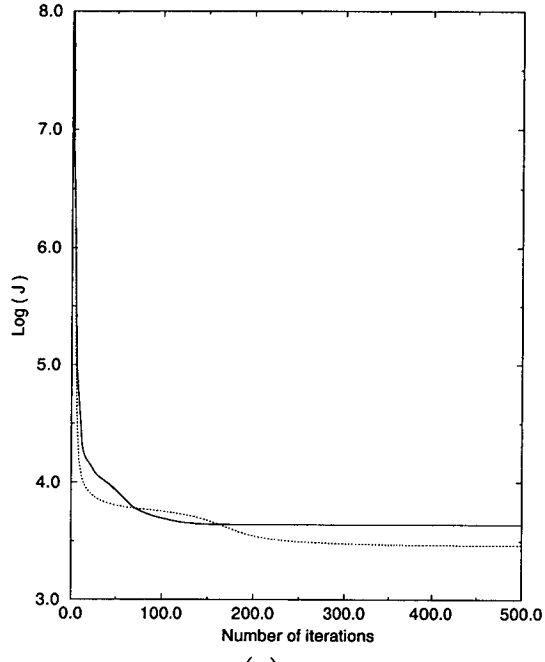
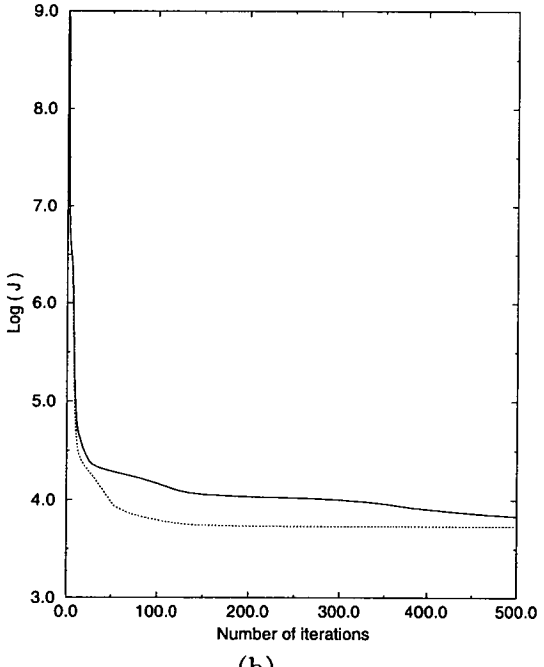


Figure 7.2: A set of the turbulent-degraded images.

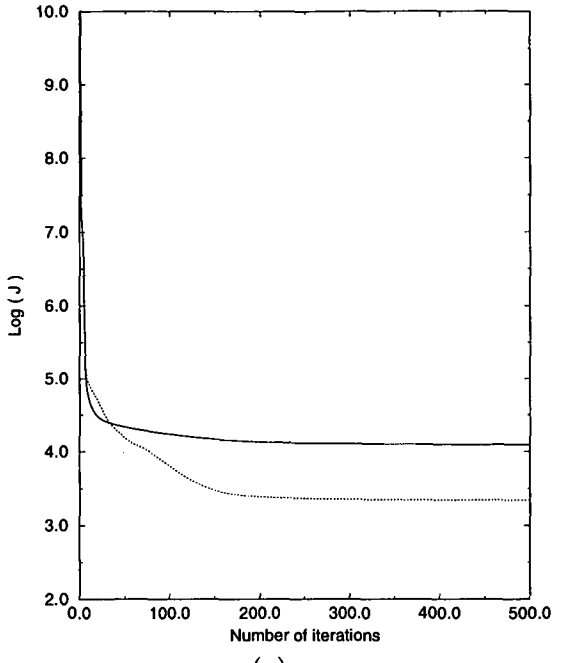


(a)

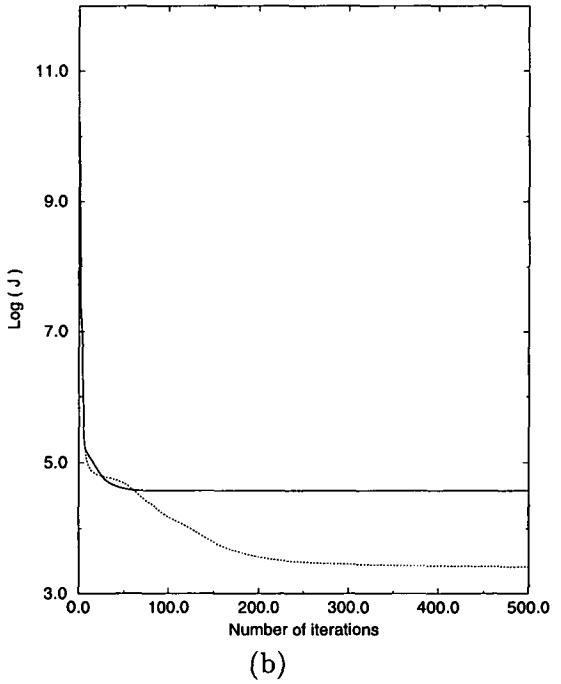


(b)

Figure 7.3: A comparison of the convergence behaviour of the original projection-based algorithm and the accelerated algorithm using momentum for (a) one frame case and (b) two frames case. The solid line is the log of the error from the original projection-based algorithm while the dotted line is that from the accelerated algorithm using momentum.



(a)



(b)

Figure 7.4: A comparison of the convergence behaviour of the original projection-based algorithm and the accelerated algorithm using momentum for (a) three frames case and (b) four frames case. The solid line is the log of the error from the original projection-based algorithm while the dotted line is that from the accelerated algorithm using momentum.

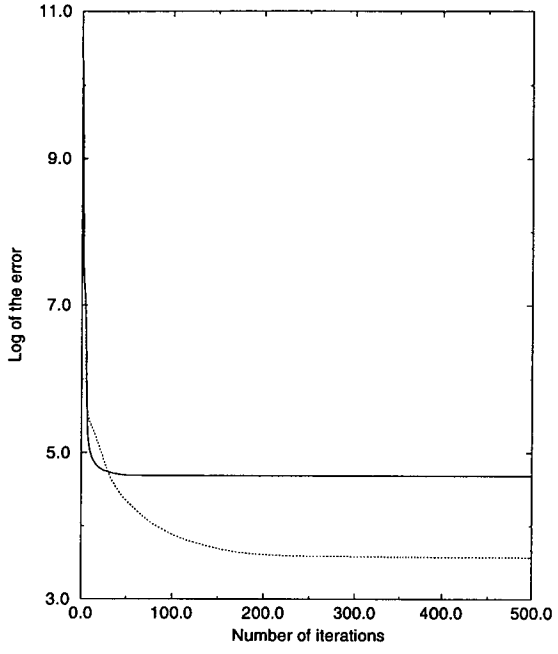


Figure 7.5: A comparison of the convergence behaviour of the original projection-based algorithm and the accelerated algorithm using momentum for five frames case. The solid line is the log of the error from the original projection-based algorithm while the dotted line is that from the accelerated algorithm using momentum.

Figure 7.3, Figure 7.4 and Figure 7.5 show graphs of the log of the value of J for the algorithms with and without momentum for different number of frames. The errors achieved by the accelerated algorithm is better than that achieved by the original algorithm, especially for the multiple frame cases. However, it should be noted that the computation time per iteration for these two algorithms are about the same. Also, it is found that when the momentum factor is in the range of 0.6 and 0.7, the results are good for all the cases. We have tried this accelerated algorithm on some other images, the range is also found to be 0.6 to 0.7 [99].

The errors achieved by the original algorithm and the minimum errors achieved by the modified algorithm for different number of frames are summarized in Table 7.1 and Table 7.2 respectively. When there are more number of frames, more constraints are applied to the algorithm and a better estimation result is expected. However as shown in Table 7.1, the error per number of frames increases with the number of frames for the original unaccelerated algorithm. The reason is that as the number of frames increases, more unknowns need to be estimated and it needs a larger number of iterations to converge to the final results. As for the accelerated algorithm using momentum, the

Number of frames	J	J/number of frames
1	4.13×10^3	4.13×10^3
2	6.84×10^3	3.42×10^3
3	1.23×10^4	4.10×10^3
4	3.75×10^4	9.38×10^3
5	4.85×10^4	9.70×10^3

Table 7.1: The minimum errors achieved by the original projection-based algorithm for different number of frames

Number of frames	momentum term	J	J/number of frames
1	0.6	2.92×10^3	2.92×10^3
2	0.6	5.02×10^3	2.51×10^3
3	0.6	2.20×10^3	7.34×10^2
4	0.6	2.60×10^3	6.50×10^2
5	0.6	3.73×10^3	7.45×10^2

Table 7.2: The minimum errors achieved by the accelerated projection-based algorithm using the momentum term for different number of frames.

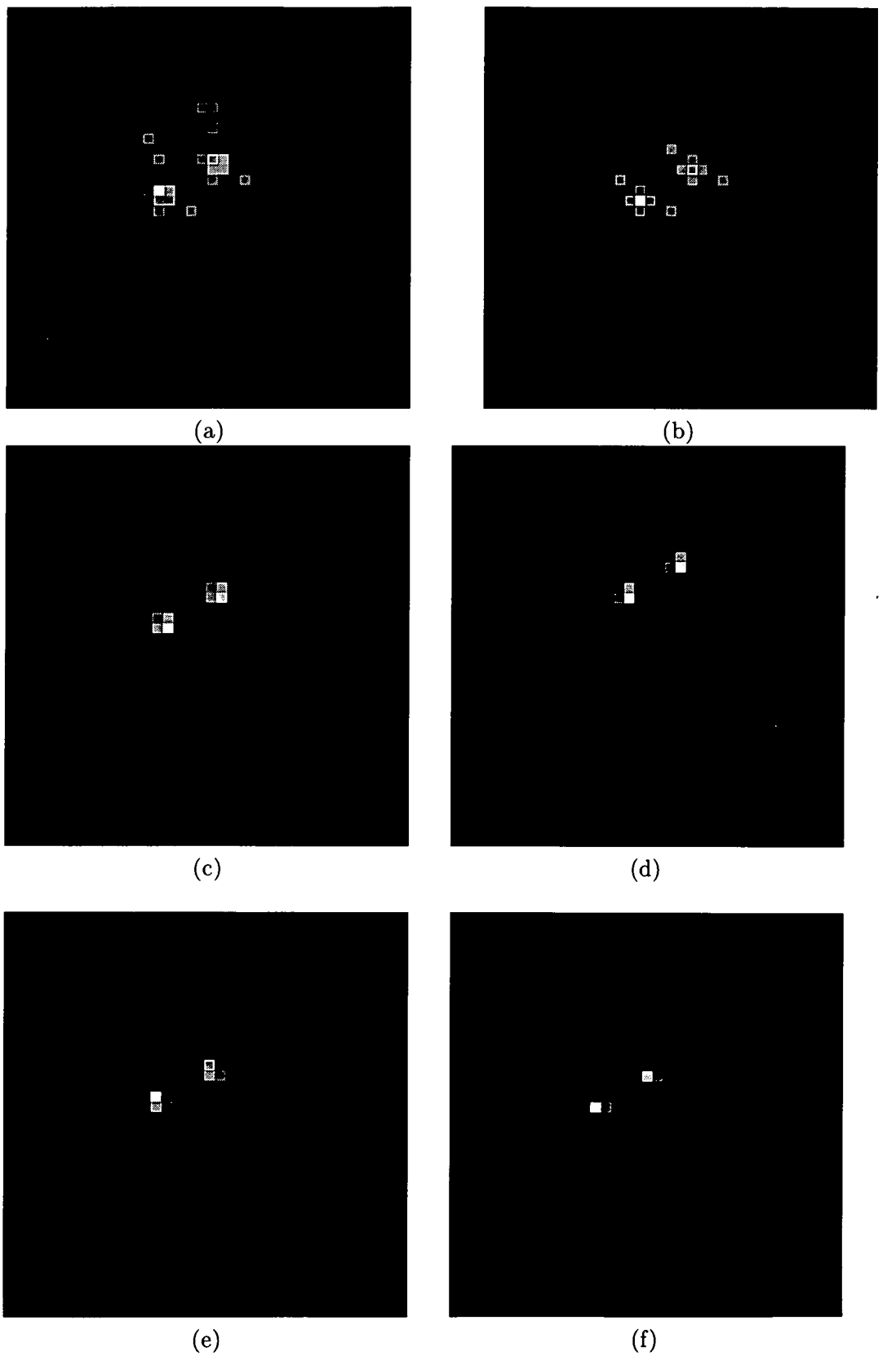


Figure 7.6: The reconstructed objects for the original and the accelerated projection based algorithms using momentum for different number of frames cases. The left column is the original algorithm while the right column is the accelerated algorithm using momentum. The first row is the results from using single frame only, the second row is the results from using two frames and the third row is that from using three frames.

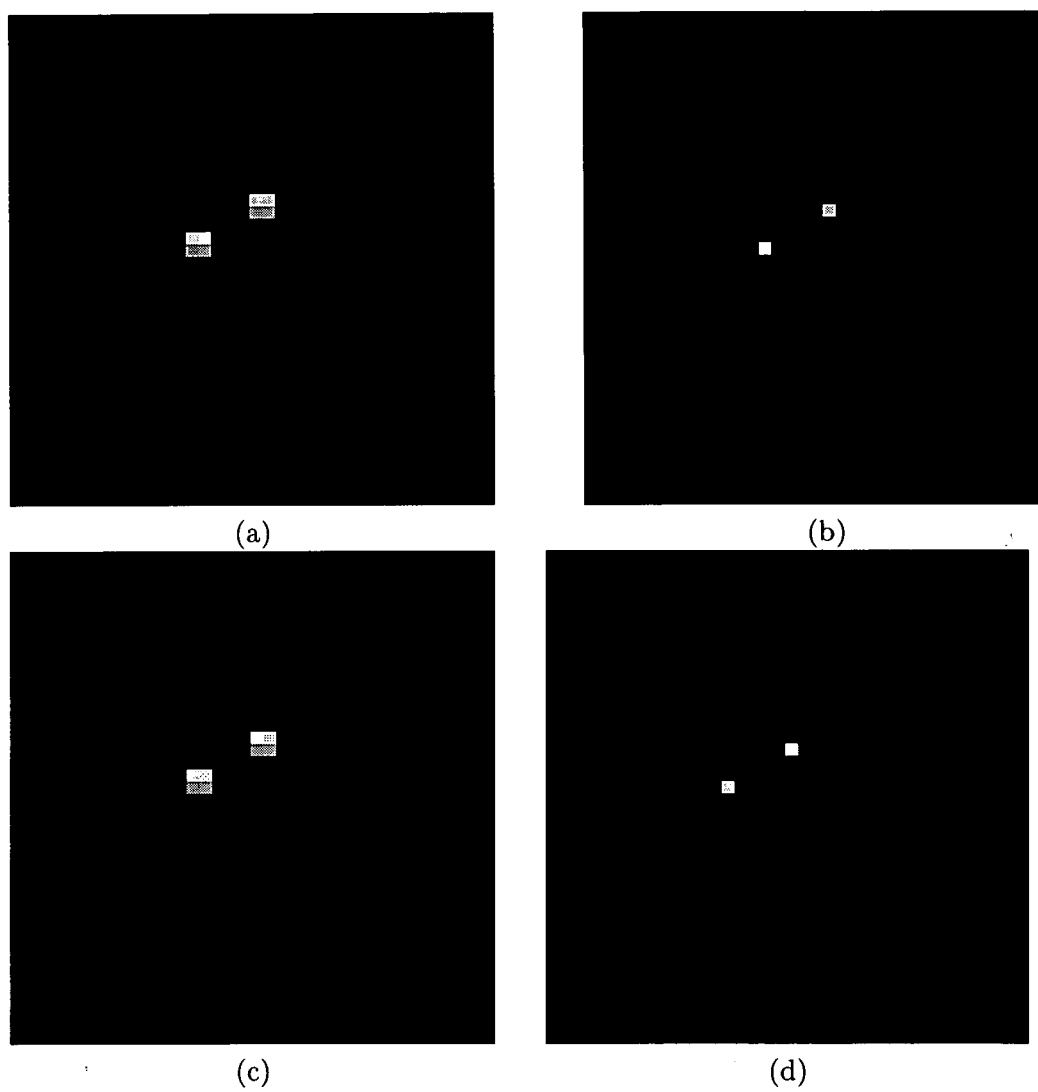


Figure 7.7: The reconstructed objects for the original and the accelerated projection-based algorithms using momentum for different number of frames cases. The left column is the original projection-based algorithm while the right column is the accelerated algorithm using momentum. The first row is the results from using four frames and the second row is that from using five frames.

Number of frames	J	J/number of frames
1	2.72×10^3	2.72×10^3
2	5.22×10^3	2.61×10^3
3	3.99×10^3	1.33×10^3
4	3.51×10^3	8.77×10^2
5	3.05×10^3	6.11×10^2

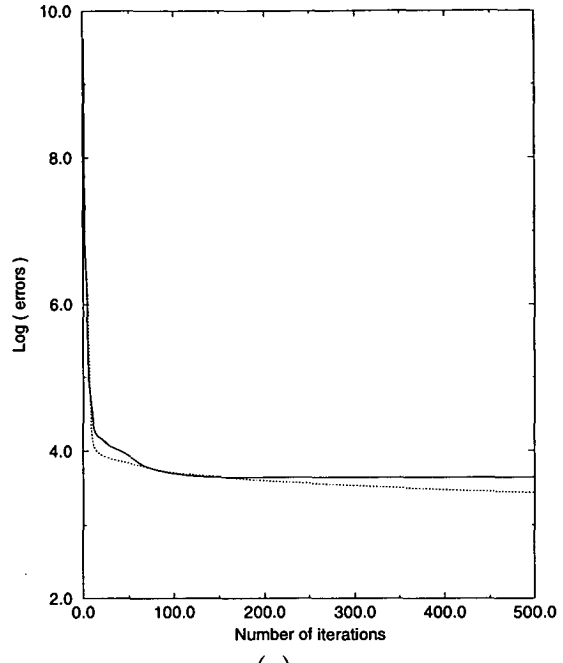
Table 7.3: The minimum errors achieved by the accelerated projection-based algorithm using the adaptive step size for different number of frames.

average error per frame decreases as the number of frames used increases. Figure 7.6 and Figure 7.7 show the reconstructed objects for using different number of frames. In general, the reconstructed object looks more similar to the original object as the number of frames increases for both the original and the accelerated algorithms.

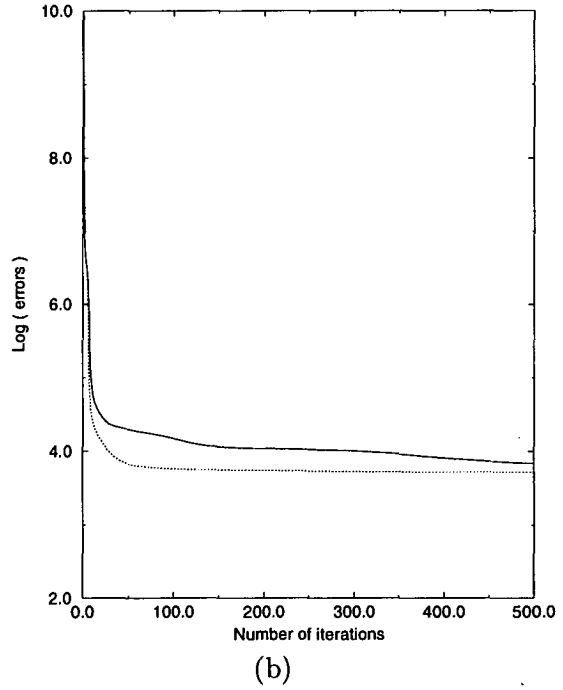
7.1.2 Use Of Adaptive Step Size

Adaptive step size approach is an ad hoc approach to speed up the convergence rate. Its idea is very simple: if the current gradient is of the same sign with the past gradient, then the step size should be larger as the solution still has not been found; however if the current gradient is of the opposite sign with the past gradient, then the solution is somewhere between the past estimate and the current estimate and thus a small step size should be chosen.

Figure 7.8, Figure 7.9 and Figure 7.10 show graphs of the log of the error J versus the number of iterations for the original algorithm and the modified algorithm using the adaptive step size. As can be seen, the adaptive step size algorithm can achieve a better convergence rate than the original algorithm. Table 7.3 summarizes the minimum errors achieved by the accelerated algorithm using the adaptive step size. As with the case of using the momentum term, the average error per frame decreases as the number of frames used increases. Figure 7.11 shows the reconstructed objects for using different

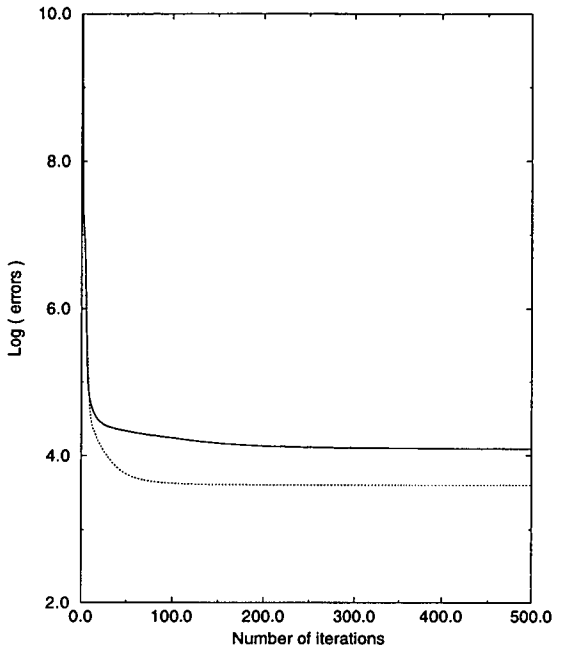


(a)

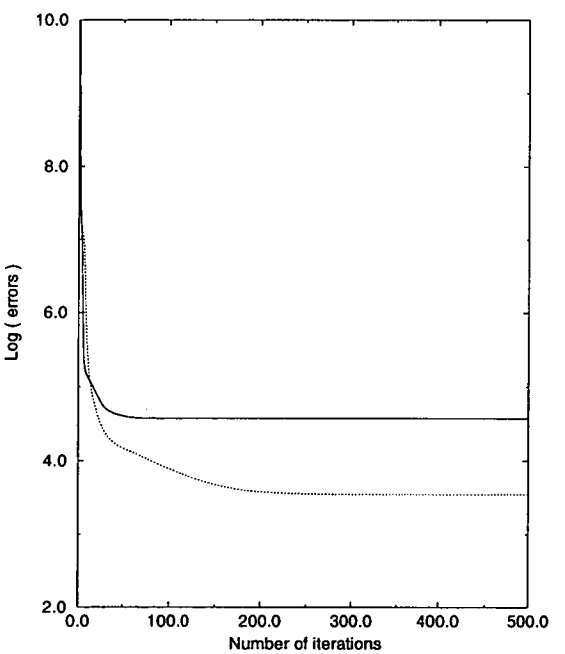


(b)

Figure 7.8: A comparison of the convergence behaviour of the original projection-based algorithm and the accelerated algorithm using the adaptive step size for (a) one frame case and (b) two frames case. The solid line is the log of the error from the original projection-based algorithm while the dotted line is that from the accelerated algorithm using the adaptive step size.



(a)



(b)

Figure 7.9: A comparison of the convergence behaviour of the original projection-based algorithm and the accelerated algorithm using adaptive step size for (a) three frames case and (b) four frames case. The solid line is the log of the error from the original projection-based algorithm while the dotted line is that from the accelerated algorithm using adaptive step size.

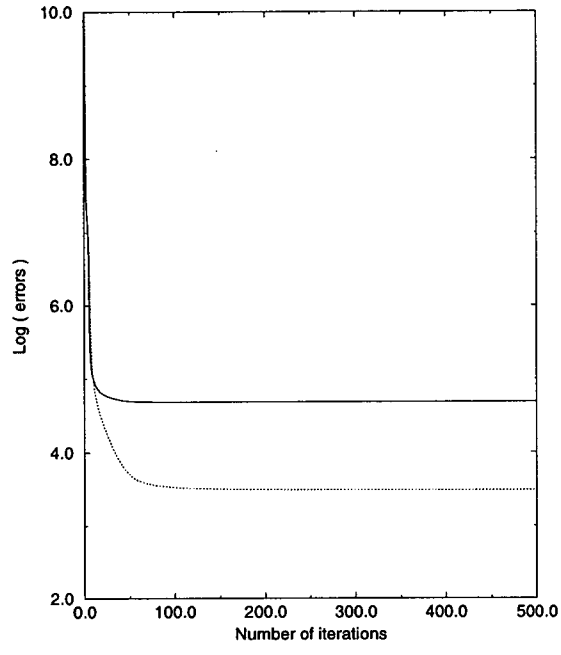


Figure 7.10: A comparison of the convergence behaviour of the original projection-based algorithm and the accelerated algorithm using adaptive step size for the five frames case. The solid line is the log of the error from the original projection-based algorithm while the dotted line is that from the accelerated algorithm using adaptive step size.

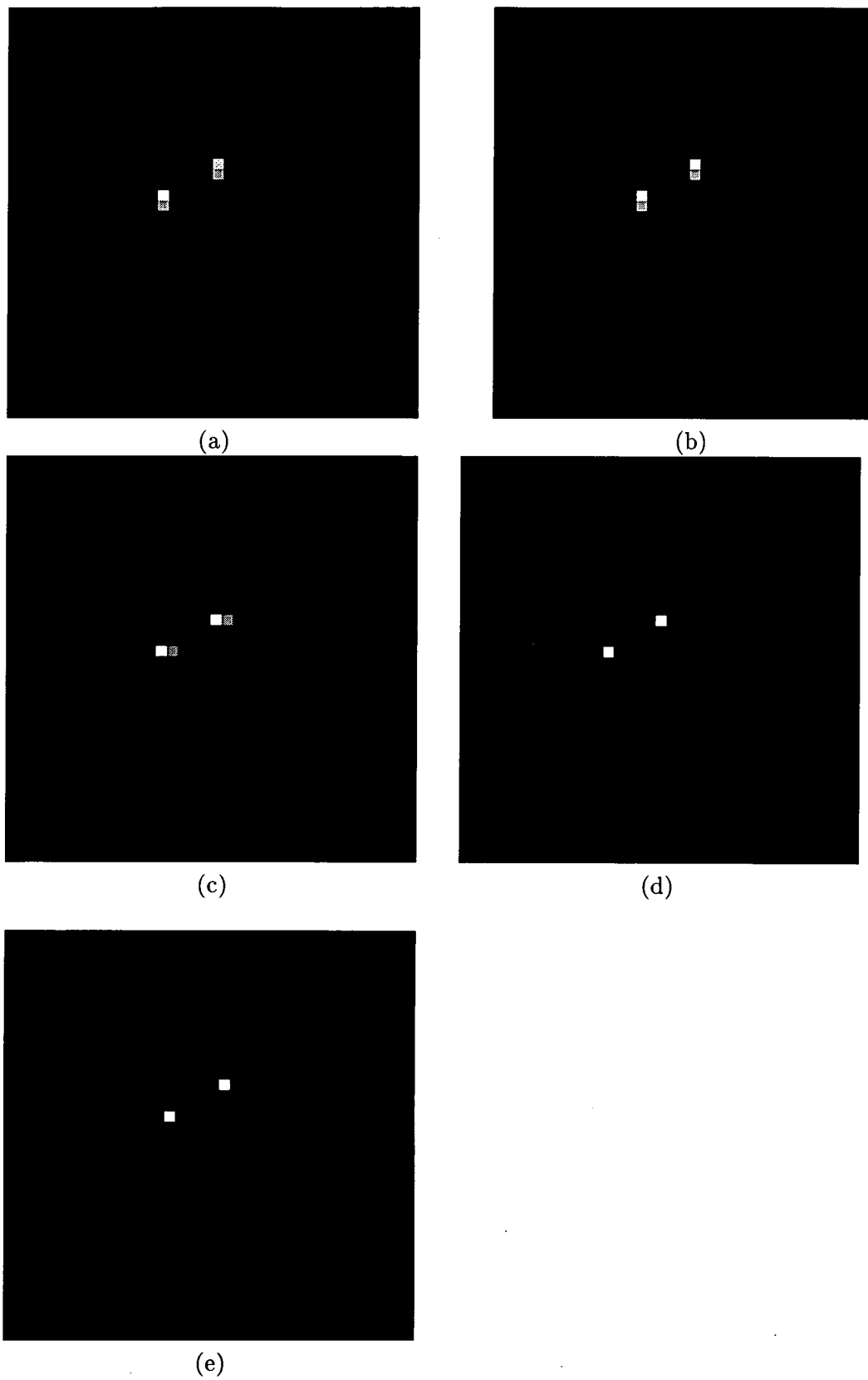


Figure 7.11: The reconstructed objects for the accelerated projection-based algorithm using adaptive step size for different number of frames, (a) 1 frame, (b) 2 frames, (c) 3 frames, (d) 4 frames and (e) 5 frames.

Number of frames	1000 iterations likelihood	1000 iterations likelihood/frame	3000 iterations likelihood	3000 iterations likelihood/frame
1	-8.82 $\times 10^4$	-8.82 $\times 10^4$	-8.82 $\times 10^4$	-8.82 $\times 10^4$
2	-1.49 $\times 10^5$	-7.46 $\times 10^4$	-1.49 $\times 10^5$	-7.46 $\times 10^4$
3	-2.18 $\times 10^5$	-7.26 $\times 10^4$	-2.18 $\times 10^5$	-7.26 $\times 10^4$
4	-2.89 $\times 10^5$	-7.23 $\times 10^4$	-2.89 $\times 10^5$	-7.23 $\times 10^4$
5	-3.52 $\times 10^5$	-7.04 $\times 10^4$	-3.52 $\times 10^5$	-7.04 $\times 10^4$

Table 7.4: The likelihood values achieved by the maximum likelihood algorithm at both 1000 and 3000 iterations.

number of frames. The reconstructed object looks more similar to the original object as the number of frames used increases.

7.2 Maximum Likelihood Approach

Maximum likelihood approach is different from the projection based algorithm. Instead of minimizing an error function, the maximum likelihood algorithm maximizes the log likelihood function, ie., it tries to find a solution that has the highest chance of making those observations. The images used are same as those used in Section 7.1. Table 7.4 summarizes the results obtained at both 1000 and 3000 iterations. As can be seen from Table 7.4, the log likelihood value per frame increases with the number of frames used. It is because more information have been incorporated into the estimation when there are more number of frames.

It is also noticed that the log likelihood values obtained at the 1000th iteration are almost same as those obtained at the 3000th iteration. It basically means that the convergence rate of the maximum likelihood algorithm is so slow that there is not much difference between the results at the 1000th iteration and that at the 3000th iteration. Figure 7.2 shows a plot of the log likelihood value with respect to the number

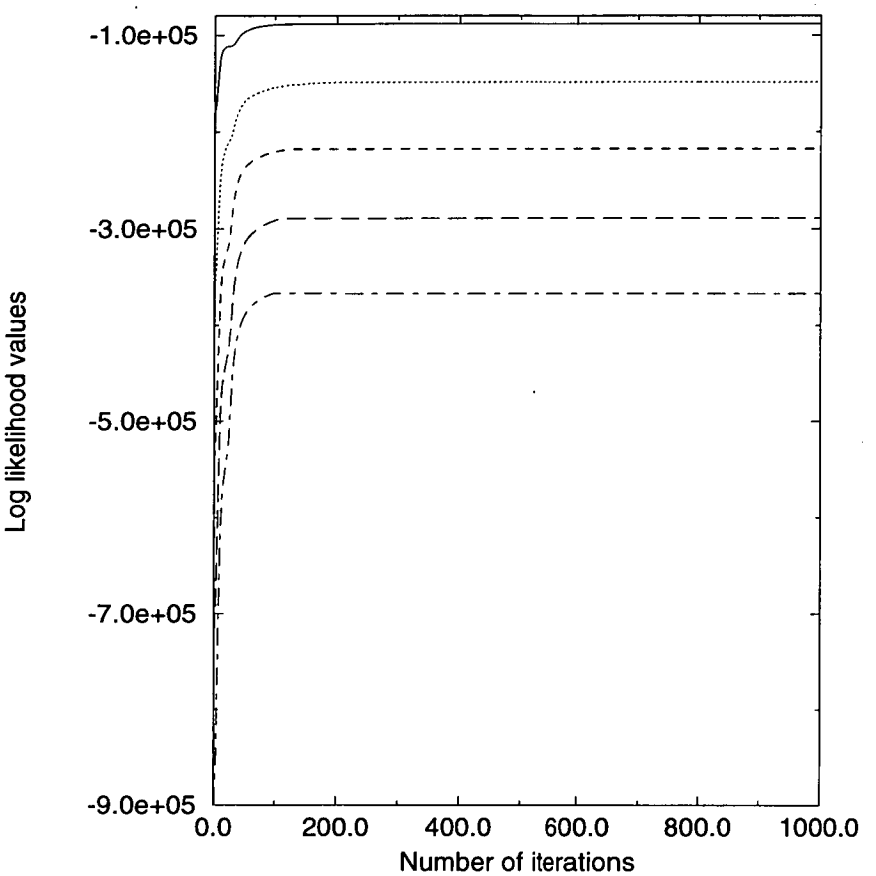


Figure 7.12: A plot of the log likelihood values versus the number of iterations for different number of frames. The solid line, the dotted line, the dashed line, the long dashed and the dot-dashed line represent 1 frame, 2 frames, 3 frames, 4 frames and 5 frames cases respectively.

of iterations for different number of frames. As is the case with the projection-based algorithm, there is generally a steep increase in the log likelihood value initially, but the value flattens out quickly and stays there for the remaining iterations. The reconstructed objects for different number of frames are shown in Figure 7.13. As more number of frames are used in the reconstruction, the reconstructed objects look more similar to the original object.

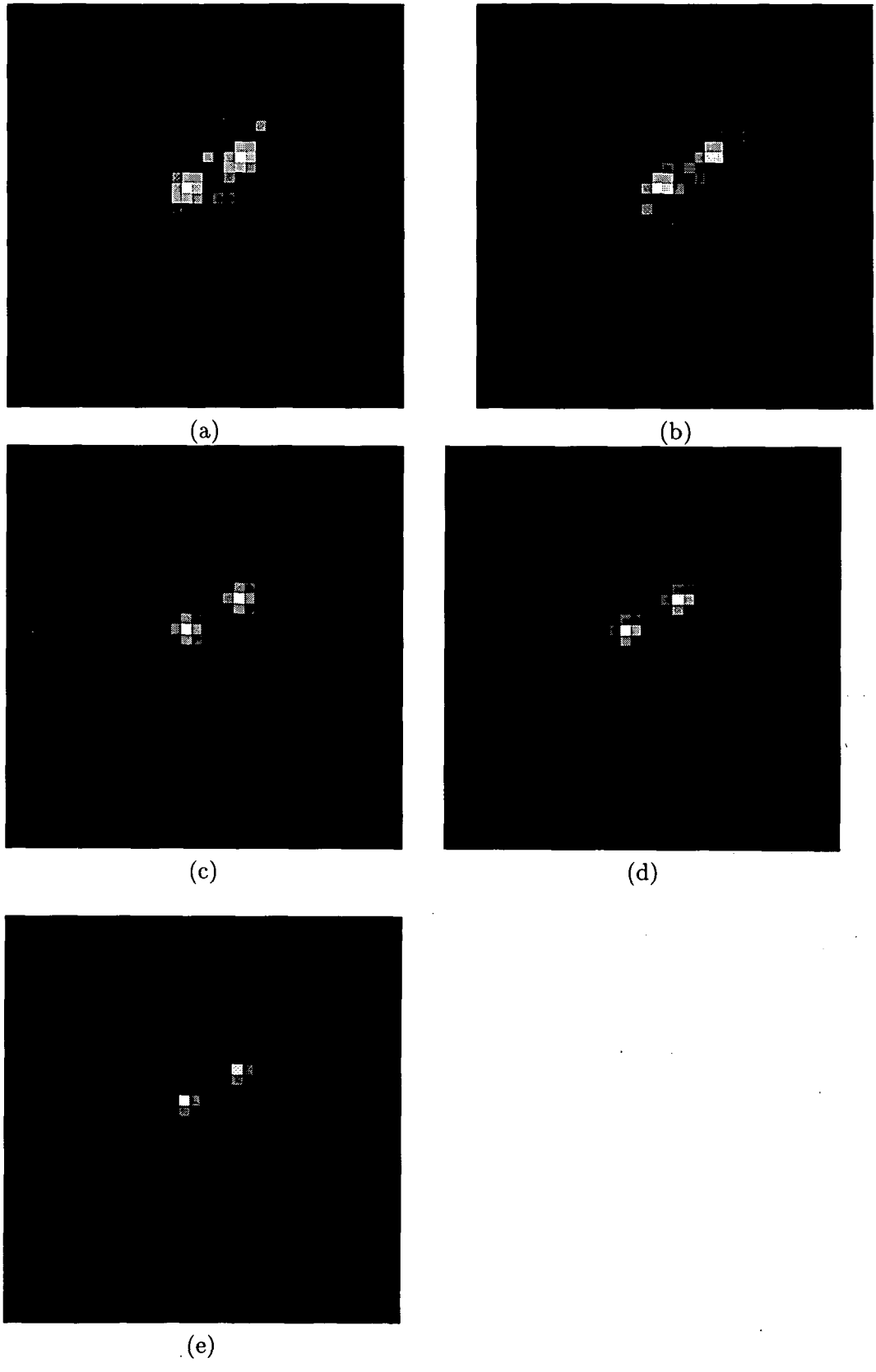


Figure 7.13: A set of the reconstructed objects for the maximum likelihood algorithm using (a) 1 frame (b) 2 frames (c) 3 frames (d) 4 frames and (e) 5 frames.

Number of frames	object error	psf error	convolution error	total error
1	8.33×10^4	5.39×10^2	1.16×10^4	9.55×10^4
2	1.28×10^5	1.70×10^3	2.54×10^4	1.55×10^5
3	1.78×10^5	4.25×10^3	3.44×10^4	2.16×10^5
4	2.30×10^5	6.42×10^3	4.77×10^4	2.85×10^5
5	2.51×10^5	7.49×10^3	5.63×10^4	3.15×10^5

Table 7.5: The object error, the point spread function error, the convolution error and the total error achieved by the conjugate gradient algorithm for different number of frames at 1000 iteration

7.3 Conjugate Gradient Approach

The conjugate gradient approach reformulates the problem so that it becomes a minimization problem as the projection based algorithm. There are basically two parts of the errors, the a priori error in which the object and the point spread function violate their a priori constraints and the convolution error. Table 7.5 and Table 7.6 summarize the performance of the conjugate gradient approach on blindly deconvolving the turbulence-degraded images at the 1000th and the 3000th iterations respectively. As can be deduced from Table 7.5, the average errors per frame are 9.55×10^4, 7.75×10^4, 7.20×10^4, 7.13×10^4, and 6.30×10^4 for 1 frame, 2 frames, 3 frames, 4frames, and 5 frames cases respectively. The average error decreases with the number of frames increases which is consistent with the results obtained from the maximum likelihood algorithm and the projection based algorithm.

There are a few interesting aspects of the conjugate gradient algorithm that can be seen from Table 7.5 and Table 7.6. The error of the point spread function is always the smallest compared to the object error or the convolution error, and the object error is always the biggest at both 1000 and 3000 iterations. It in turn means that there is an unequal weighting between the object error and the point spread function error. The conjugate gradient algorithm pays more attention on reducing the point spread function error than on reducing the object error. Table 7.7 summarizes the percentage

Number of frames	object error	psf error	convolution error	total error
1	3.79×10^4	5.81×10^1	2.90×10^3	4.09×10^4
2	5.94×10^4	2.54×10^2	7.41×10^3	6.71×10^4
3	7.56×10^4	6.07×10^2	1.24×10^4	8.86×10^4
4	1.11×10^5	1.17×10^3	1.96×10^4	1.32×10^5
5	1.14×10^5	1.46×10^3	2.43×10^4	1.40×10^5

Table 7.6: The object error, the point spread function error, the convolution error and the total error achieved by the conjugate gradient algorithm for different number of frames at 3000 iteration

changes between the errors achieved at 1000 and 3000 iterations. The point spread function error can always be reduced to a greater extent than the object error. This unequal weighting problem between the a priori error in the object and the point spread function remains to be solved. The reconstructed objects using the conjugate gradient algorithm for different number of frames are shown in Figure 7.14.

Another unequal weighting behaviour occurs between the a priori error and the convolution error. This unequal weighting, however, can be used to advantage when dealing with inherently variable nature of actual data as has been discussed in Section 6.2. When the observed convolution is scaled by ζ, the components increase only by $\sqrt{\zeta}$. The new error function to be minimized is then given in Equation (6.15) and repeats here as,

$$E_t^{new} = \zeta^2 E_c + \zeta E_i \tag{7.1}$$

Hence one can control the relative importance between the convolution error and the a priori error by simply scaling the observed convolution. If ζ is chosen to be smaller than 1, then

$$\zeta^2 < \zeta \tag{7.2}$$

the minimization process would put more emphasis in reducing the a priori error. It in turn means that the support and the positivity constraints are applied more strictly.

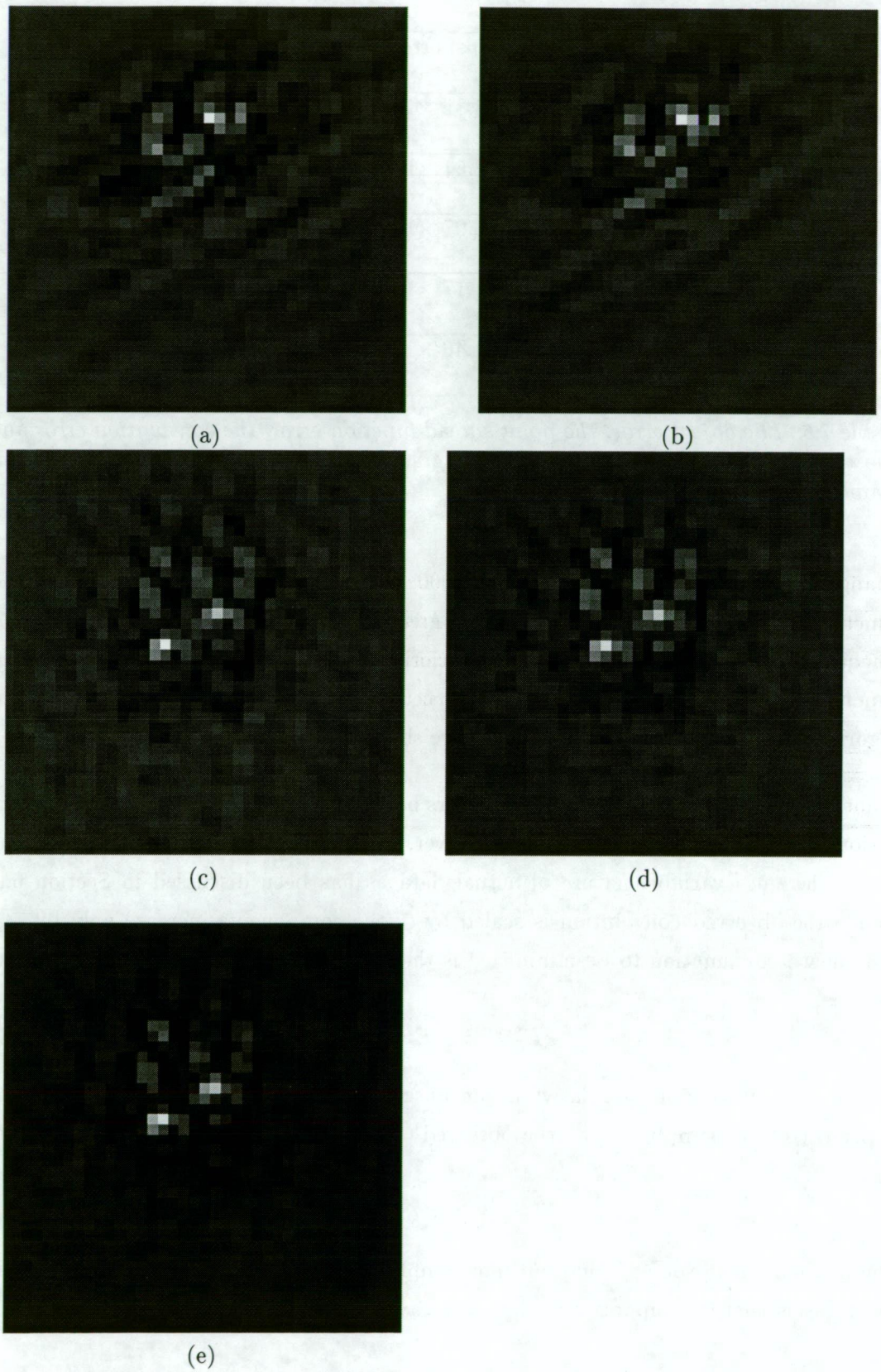


Figure 7.14: The reconstructed objects for the conjugate gradient algorithm at 1000 iterations using (a) 1 frame (b) 2 frames (c) 3 frames (d) 4 frames (e) 5 frames.

Number of frames	object error	psf error	convolution error
1	54.5%	89.2%	75.0%
2	53.6%	85.1%	70.8%
3	57.5%	85.7%	64.0%
4	51.7%	81.8%	56.8%
5	54.6%	80.5%	58.9%

Table 7.7: The percentage difference between the object error, the point spread function error and the convolution error achieved at 1000 and 3000 iterations for the conjugate gradient algorithm

In order to demonstrate this point, the algorithm was tested using different values of ζ. Table (7.8) and Table (7.9) summarize the resultant errors and the percentage errors for scaling down the observed convolution by 10, 100 and 1000 at the 100th iteration respectively. It should be noted that the region of support is the same despite the value of ζ.

As shown in Table (7.9), when ζ is large, the total a priori error which include the object error and the point spread function error is large compared with the convolution error. It means that the algorithm puts more emphasis in reducing the convolution error. The a priori constraint, that is the support and the positivity constraints, is applied loosely only. Hence the reconstructed objects are not clear and have negative valued pixels.

But when ζ is small, the a priori error decreases significantly compared with the convolution error. It is because the algorithm puts more emphasis in reducing the a priori error. In other words, the a priori constraint is applied more strictly when ζ is large. Especially when $\zeta = 0.001$, most of the errors are from the convolution error. In this case, most of the pixels of the reconstructed object would satisfy the positivity and the support constraints. The reconstructed objects for the case with $\zeta = 0.001$ are shown in Figure 7.15.

It should be noted that the results obtained are based on a set of five differently initial-

number of frames	errors	ζ 0.1	ζ 0.01	ζ 0.001
1	total	2.28×10^{1}	1.81×10^{-1}	1.38×10^{-2}
	object	3.34	2.57×10^{-3}	2.14×10^{-5}
	psf	2.04×10^{-1}	1.15×10^{-3}	1.92×10^{-5}
	conv	1.93×10^{1}	1.77×10^{-1}	1.38×10^{-2}
2	total	4.88×10^{1}	1.65×10^{-1}	9.50×10^{-3}
	object	1.18×10^{1}	4.09×10^{-3}	1.26×10^{-5}
	psf	1.55×10^{-1}	1.54×10^{-3}	8.20×10^{-6}
	conv	3.69×10^{1}	1.59×10^{-1}	9.48×10^{-3}
3	total	2.61×10^{3}	1.12	1.98×10^{-2}
	object	1.72×10^{3}	8.43×10^{-3}	1.55×10^{-5}
	psf	1.00×10^{1}	6.14×10^{-3}	1.85×10^{-5}
	conv	8.78×10^{2}	1.10	1.97×10^{-2}
4	total	2.84×10^{3}	3.94	4.43×10^{-2}
	object	2.08×10^{3}	9.47×10^{-3}	1.94×10^{-5}
	psf	8.05	6.63×10^{-2}	4.71×10^{-5}
	conv	7.60×10^{2}	3.87	4.42×10^{-2}
5	total	4.48×10^{3}	4.63	4.81×10^{-2}
	object	2.74×10^{3}	3.22×10^{-2}	3.27×10^{-5}
	psf	1.97×10^{1}	4.68×10^{-2}	6.48×10^{-5}
	conv	1.72×10^{3}	4.55	4.80×10^{-2}

Table 7.8: The object error, the point spread function error and the convolution error achieved by the conjugate gradient algorithm for different values of ζ.

Number of frame	Percentage errors	ζ 1	ζ 0.1	ζ 0.01	ζ 0.001
1	object	87.2	14.6	1.4	0.15
	psf	0.60	0.90	0.64	0.14
	conv	12.2	84.6	97.9	99.7
2	object	82.6	24.1	2.48	0.13
	psf	1.10	0.32	0.93	0.09
	conv	16.4	75.6	99.6	99.8
3	object	82.4	65.9	0.75	0.08
	psf	1.97	0.39	0.55	0.09
	conv	15.9	33.7	98.7	98.8
4	object	80.7	73.0	0.24	0.04
	psf	2.25	0.28	1.67	0.11
	conv	16.7	26.7	97.3	99.8
5	object	79.7	61.6	0.70	0.07
	psf	2.38	0.44	1.01	0.13
	conv	17.9	38.5	98.3	99.8

Table 7.9: The percentage errors in the object, the point spread function and the convolution error with respect to the total area for different values of ζ.

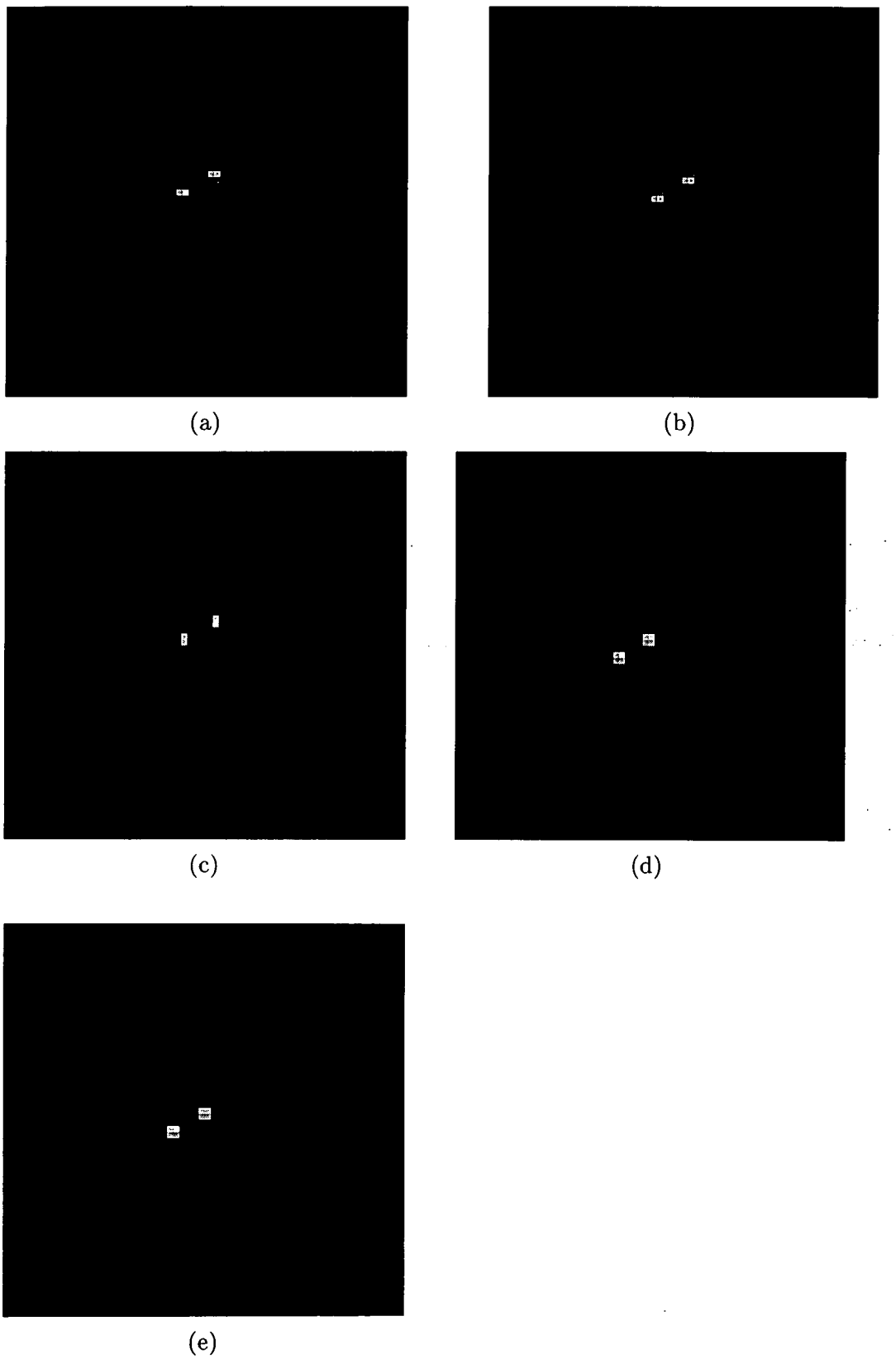


Figure 7.15: The reconstructed objects for the conjugate gradient algorithm at 1000 iterations using (a) 1 frame (b) 2 frames (c) 3 frames (d) 4 frames (e) 5 frames with $\zeta = 0.001$.

Algorithms	Maximum Likelihood	Conjugate Gradient	Projection
Noise Statistics	Poisson distributed	Gaussian	Gaussian
A priori knowledge	exact	inexact	exact

Table 7.10: The differences between the three blind deconvolution algorithms on the underlying noise statistics and the application of the a priori knowledge.

ized lowpass filtered white noise. The use of lowpass filtered white noise is because we assume that we do not have any a priori knowledge concerning the shape/distribution of the object and the psf and we want to see how the algorithm would perform under this general case. However, it is possible to include some a priori knowledge into the initial estimates as discussed later in Section 7.4.3.

7.4 Comparison Of Different Blind Deconvolution Algorithms

In this section, a comparison of the three blind deconvolution algorithms described in Chapter 6 is presented. In summary, all these three algorithms can be formulated through the Bayes theorem but with different assumption on the noise statistics and different application of the a priori information. Table 7.10 summarises these differences. The comparison of these three algorithms is based on the convergence rate and the quality of the reconstruction. As these three algorithms have different objective functions, ie., maximum likelihood maximizes the log likelihood function, conjugate gradient and the projection based minimizes an error function, it is better to use some other metric to quantify the performance of the algorithms. In Peng and Stark's paper [28], they suggested that a possible way is to use the similarity metric. It is defined as the correlation between the original object and the reconstructed object normalized by their respective norms,

$$S = \max_{x_0,y_0} \frac{\int\int (f(x,y)\hat{f}(x-x_0,y-y_0))dxdy}{||f||\,||\hat{f}||} \tag{7.3}$$

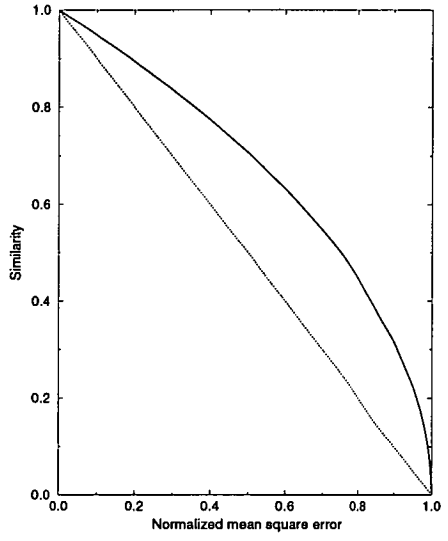


Figure 7.16: A Plot of the similarity metric against the normalized mean square error. The dotted line is a 45 degrees straight line.

Yang et al [91] claim that the similarity metric S is a measure of the similarity of the shapes of f and $\hat{f}$. Therefore, it is more closely correlated to the visual quality of the reconstruction than the mean-square distance.

Another possible metric used to quantify the performance of the algorithm is the normalized mean square error suggested by Miura et al [100]. It is defined as,

$$mse = \min_{x_0,y_0,\alpha} \frac{\int\int (f(x,y) - \alpha\hat{f}(x - x_0,, y - y_0))^2 dxdy}{\int\int (f(x,y))^2 dxdy} \tag{7.4}$$

The normalized mean square error can be used to provide an average error for the estimation. In fact, it is easy to show that S and mse is related by

$$mse = 1 - S^2 \tag{7.5}$$

A Plot of S against mse is shown in Fig 7.4. If the estimation is of good quality, then mse provides more discriminating power than S. Conversely, S should be used when the quality of the reconstructed image is poor. Both measures are used in the simulation results.

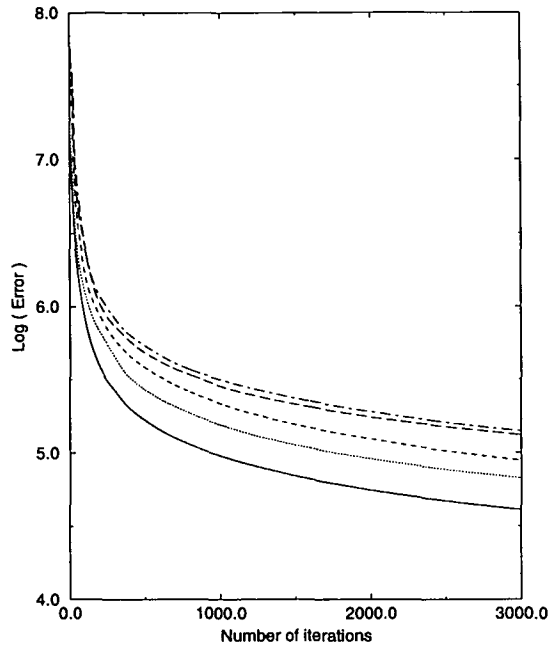


Figure 7.17: The convergence behaviours of the conjugate gradient algorithm. The solid line, the dotted line, the dashed line, the long dashed line and the dot-dashed line represent one frame, two frame, three frame, four frame and five frame cases respectively.

7.4.1 Convergence Rate

First of all, the comparison is made on the convergence rate of the three algorithms. The convergence behaviour of the projection based algorithm and the maximum likelihood algorithm have been shown in Figures 7.3, 7.4 and Figure 7.2 respectively. The convergence behaviour of the conjugate gradient algorithm is shown in Figure 7.17. It should be noted that all these three graphs are plotted in logarithmic scale. All curves show a steep change initially, ie., rapidly reduced the error or increase the log likelihood value by several orders of magnitude, but flatten out afterward. For the projection-based and the maximum likelihood algorithms, there are no noticeable changes after a certain number of iterations. But for the conjugate gradient algorithm, the error keeps decreasing although the rate of decrease is reduced with the number of iterations increases.

In order to compare their convergence behaviour, two metrics are used. The first one is the numbers of iterations for the curves to flatten out and the second one is the numbers of iterations taken to reach 90% of the final value. Table 7.11 summarizes these values.

algorithm	Number of frames	number to flatten out	number to reach 90% value
conjugate gradient algorithm	1	3000	2609
	2	3000	2641
	3	3000	2651
	4	3000	2666
	5	3000	2704
projection based algorithm	1	250	232
	2	150	114
	3	175	155
	4	350	309
	5	250	200
maximum likelihood algorithm	1	100	51
	2	150	61
	3	125	64
	4	100	69
	5	100	70

Table 7.11: A comparison of the convergence behaviour of the conjugate gradient, projection-based and the maximum likelihood algorithms. These three algorithms are compared in terms of the number of iterations taken to reach the ninety percent of the final value and the total number of iterations that has been taken to a stage where there is no obvious changes in the final value.

As shown in Table 7.11, the maximum likelihood algorithm takes the least number of iterations to both the final value and the 90% of the final value while the conjugate gradient algorithm takes the greatest number of iterations to achieve that.

7.4.2 Quality of Estimation in terms of Similarity and Mean Square Error

The reconstructed objects have been shown in Figures 7.6, 7.7 Figure 7.13 and Figure 7.14 for projection-based, maximum likelihood and the conjugate gradient algorithms respectively. Both similarity metric and the mean square error are computed and they are summarized in Table 7.12. As shown in Table 7.12, the projection-based algorithm

algorithm	Number of frames	similarity	mean square error
conjugate gradient algorithm	1	0.4617	0.7868
	2	0.5189	0.7308
	3	0.5345	0.7143
	4	0.5468	0.7010
	5	0.5946	0.6465
projection based algorithm	1	0.8650	0.2518
	2	0.8490	0.2792
	3	0.9951	0.0098
	4	0.9984	0.0032
	5	0.9992	0.0016
maximum likelihood algorithm	1	0.6310	0.6018
	2	0.5811	0.6623
	3	0.7764	0.3972
	4	0.8266	0.3167
	5	0.8047	0.3525

Table 7.12: A comparison of the conjugate gradient, projection-based and the maximum likelihood algorithms in terms of the quality of the reconstructed images. These three algorithms are compared in terms of the similarity and the mean square error.

can achieve the best result, in terms of both the similarity metric and the normalized mean square error.

7.4.3 A priori constraints

Generally, the a priori constraints have been applied absolutely in the projection based and the maximum likelihood algorithms, ie., there are no negative estimated pixels in the estimated objects and all the pixel values outside the support region are zero. Hence the estimated objects for these two algorithms all look clear.

However, for the conjugate gradient algorithm, the a priori constraints are not applied absolutely. Hence the reconstructed objects can have negative pixel values and can have values outside the support region. It in turns means that the estimated object is not as clear as the one obtained from the projection based or the maximum likelihood algorithms. Despite that, the reconstructed objects can still be recognised.

It should be emphasised that whether a priori constraints can be applied absolutely will depend on the real situation, but it need not necessarily be the case with real data. Hence conjugate gradient algorithm is in this way more flexible in which it does not seek to meet the constraints absolutely, but to find a solution that compromises between the error resulting from violating the a priori constraints and the convolution error.

It is also possible to include some a priori knowledge into the initial estimates. As the error to be minimized in the blind deconvolution problem is not quadratic, there exists local minima in addition to the global minimum. If some a priori infomration is included in the form of the starting point, the initial estimate is closer to the required minimum. Hnece, there is a greater chance of converging to the required solution and the convergence rate is expected to be more rapid. This is in analogue to the multigrid/relaxation method used in computer vision. An estimate is first generated at lower resolution and then this estimate is sued as a starting point for the processing at higher resolution. In other words, the process can be seen as providing some a priori information (from the lower resolution) into the processing at the higher resolution. This point will be further discussed in Section 8.2 and in paper [101].

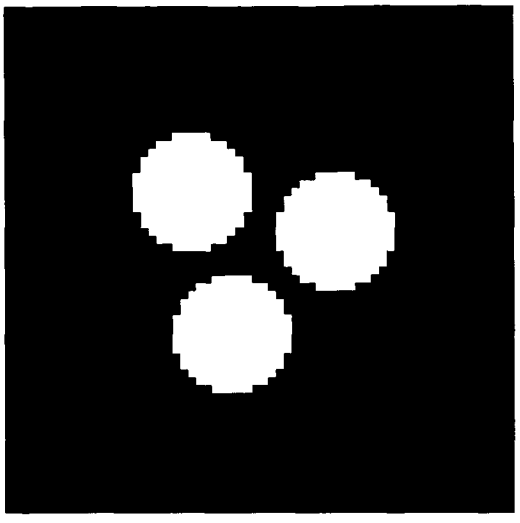


Figure 7.18: The object used to investigate the noise performance of the three blind deconvolution algorithms.

7.5 Noise Performance

Images are often corrupted by noise in the practical situation. The dominant form of noise in astronomical imaging is photon noise. In this section, the performance of the three algorithms in the presence of noise is investigated. The original object used is shown in Figure 7.18. It is a 64 × 64 image of a triple star modelled by three circles [91]. It is blurred by a truncated Gaussian function defined as,

$$h(x,y) = \begin{cases} \frac{1}{2\pi\sigma^2}\exp[-\frac{x^2+y^2}{2\sigma^2}], & x^2+y^2 \leq 4\sigma^2 \\ 0 & otherwise \end{cases} \tag{7.6}$$

where $\sigma = 7$. The blurring function is shown in Figure 7.19. The resultant blurred image is shown in Fig 7.20. Three levels of noise were examined. For the first, the total number of photons in the observed image was assumed to be 10^5, for the second 5×10^5 and for the third 10^6. The resultant images for noise levels 1, 2 and 3 are shown in Figure 7.21. These three noise levels correspond to a signal to noise ratio (SNR) of 18 db, 23 db and 28 db respectively.

It should be noted this example is likely to be difficult for any blind deconvolution algorithm because of the choice of the point spread function. If for the moment we neglect the effect of truncation, then the point spread function can be represented as a

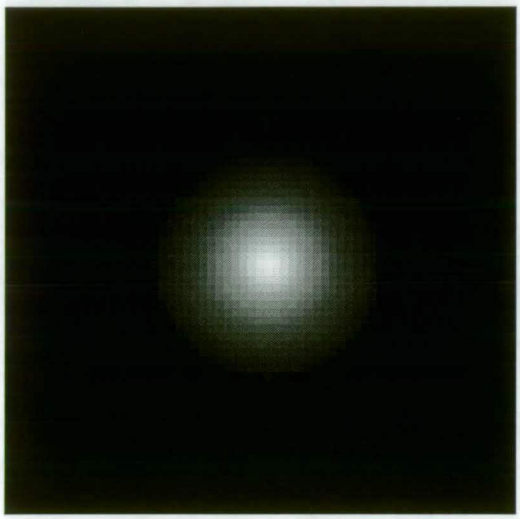


Figure 7.19: The truncated Gaussian function used to blur the object shown in Fig 7.18.

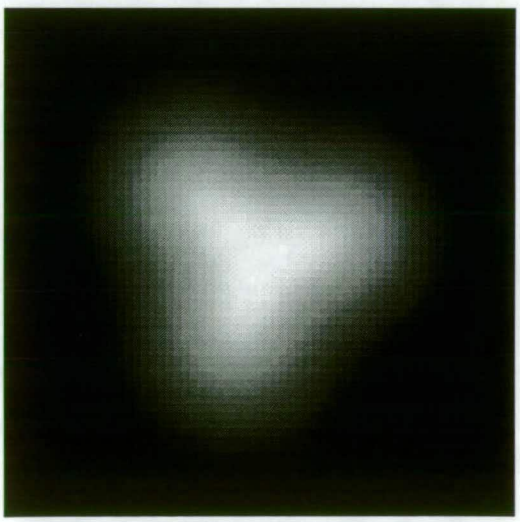
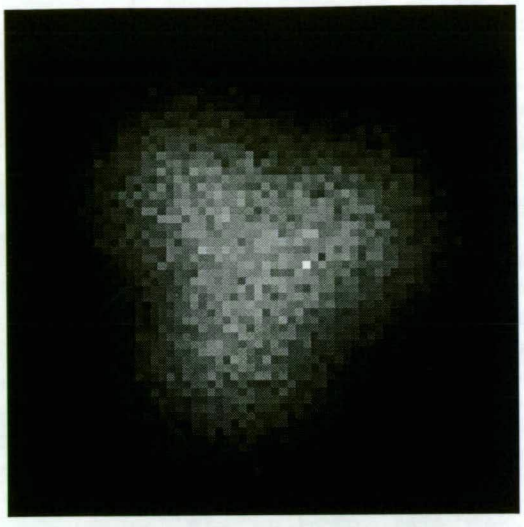
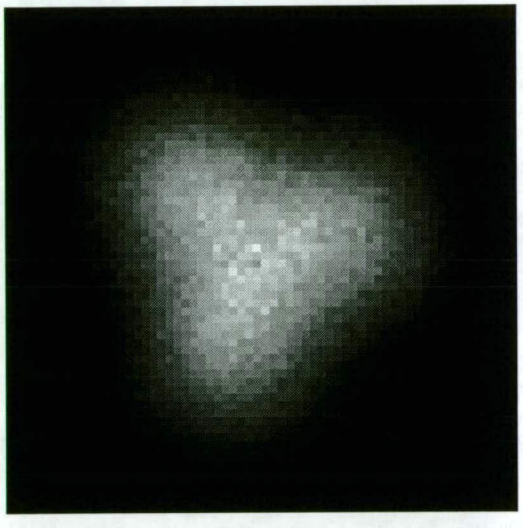


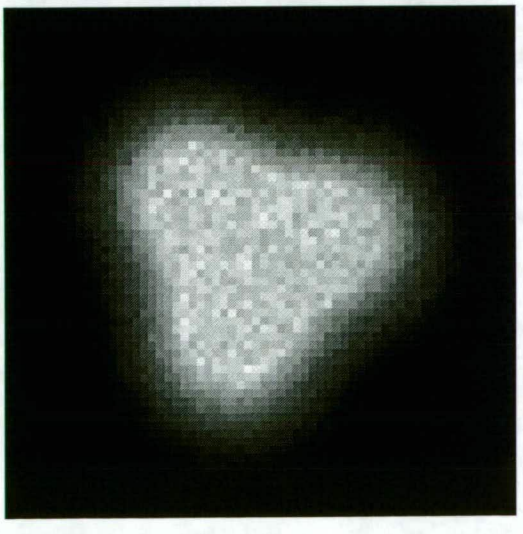
Figure 7.20: The resultant blurred image formed by convolving Fig 7.18 and Fig 7.19.



(a)



(b)



(c)

Figure 7.21: Photon-limited data : (a) 10^5 counts, (b) 5×10^5 counts and (c) 10^6 counts.

$\hat{f}(x,y)$	$S(f,\hat{f})$	$\hat{h}(x,y)$	$S(h,\hat{h})$
Figure 7.22a	0.895	Figure 7.22b	0.949
Figure 7.23a	0.075	Figure 7.23b	0.985

Table 7.13: A comparison of the similarity metrics for the different estimates of components for the blind deconvolution problem, shown in Figure 7.20.

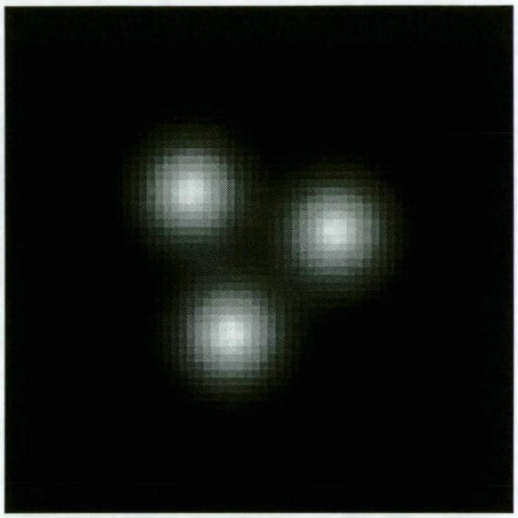
convolution of any two Gaussian of sizes defined by σ_1 and σ_2 provided

$$\sigma_1{}^2 + \sigma_2{}^2 = \sigma^2 \tag{7.7}$$

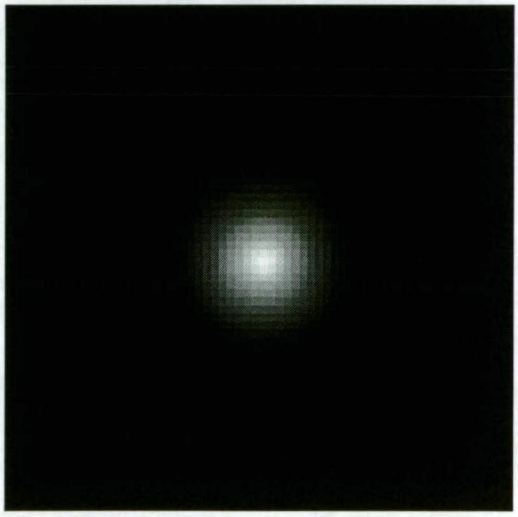
The original true image can also be approximately represented by the convolution of a circle with three delta functions. It is thus a relatively simple matter to generate a large number of solutions which almost exactly convolve to the observed convolution, for example Figures (7.22) and (7.23). The convolution of the components shown in Figure (7.22) has a similarity measure of $S = 0.997$ when compared with the original convolution. This indicates that there is very little difference between this and the original convolution.

Despite the extremely close correlation between the convolution of the components in Figures (7.20), (7.22) and (7.23), there is a significant deviation in the similarity metrics for the components as shown in Table (7.13). The solutions shown in Figures (7.22) and (7.23) do differ in the size for their supports, and hence an algorithm must place very high emphasis on the support constraint in order to distinguish between these solutions. It might be suggested that in the more realistic situations of the point spread function being an untruncated Gaussian, an exact support constraint would be unavailable, whereupon overestimation of the size of the image support would render discrimination between the ambiguous solutions impossible.

We investigate the blind deconvolution of Figure (7.21) from a range of starting points formed by lowpass filtering white noise. The conjugate gradient method was started with the energy within the estimated supports of the point spread function and image normalized to the same value (one in the simulations below). A higher weight was also placed on the support constraint by scaling down the convolution. For both the

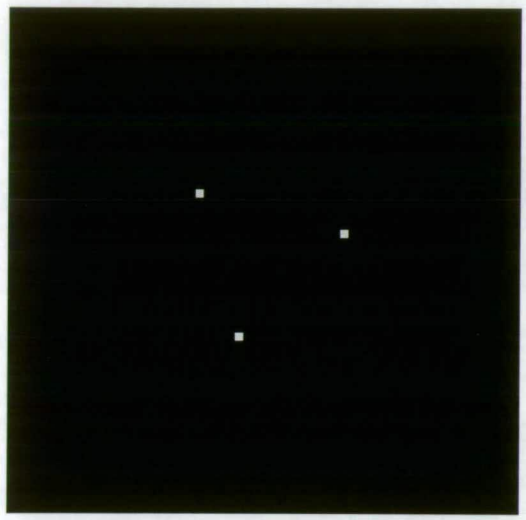


(a)

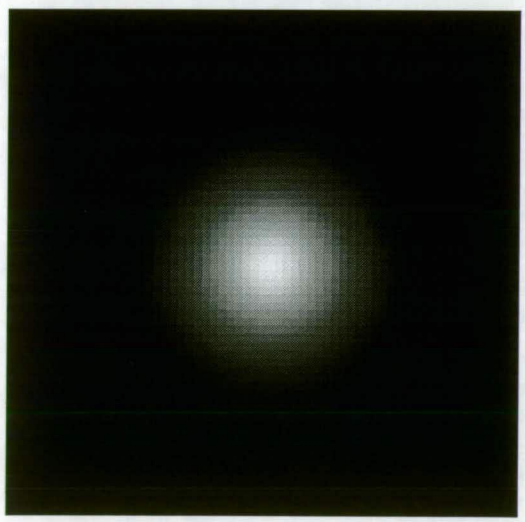


(b)

Figure 7.22: Potential ambiguous solution to the blind deconvolution of Figure 7.20. The convolution of (a) $\hat{f}(x,y)$ and (b) $\hat{h}(x,y)$ is almost visually indistinguishable from Figure 7.20.



(a)



(b)

Figure 7.23: Another ambiguous solution to the blind deconvolution of Figure 7.20. The convolution of (a) $\hat{f}(x,y)$ and (b) $\hat{h}(x,y)$ is again visually indistinguishable from Figure 7.20.

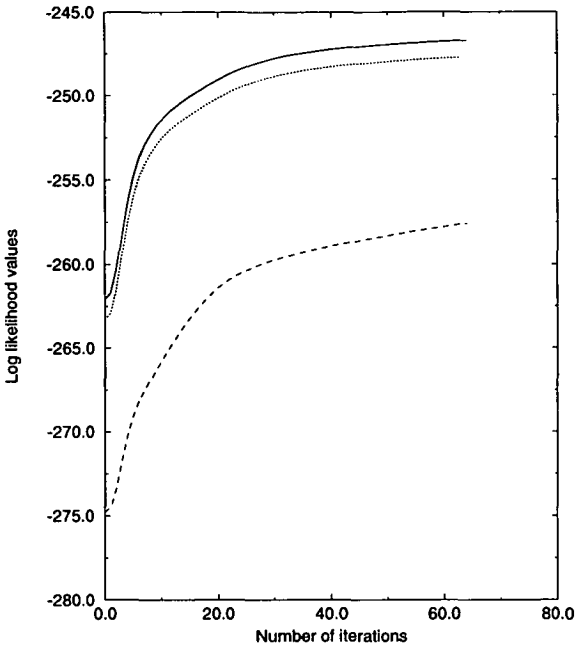


Figure 7.24: The performance of the maximum likelihood algorithm in the photon-limited case. The solid line, the dotted line and the dashed line represents 10^5, 5×10^5 *and* 10^6 *photon counts respectively.*

maximum likelihood and the projection-based algorithms, they were started with all ones within the estimated supports of the image and all zeros for the point spread function.

The plots of the log likelihood values for the maximum likelihood algorithm, the log of the errors for the conjugate gradient algorithm and that for the projection-based algorithm are shown in Figure 7.24, Figure 7.25 and Figure 7.26 respectively. The curves are monotonically increasing for the maximum likelihood algorithm and are monotonically decreasing for the other two algorithms, although there is a significant slowing in the convergence rate after a certain number of iterations. By contrast, the similarity metric for these three algorithm shown in Figure 7.27, Figure 7.28 and Figure 7.29 have a different performance. The curves in general peak and then decrease monotonically. A similar behaviour is observed for the normalized mean square error as shown in Figure 7.30, Figure 7.31 and Figure 7.32. It first decreases to a minimum value and then increases monotonically. Table 7.14 summarizes the peak similarity metrics for the three different algorithms.

Therefore, increases in log likelihood values does not imply that the similarity metric

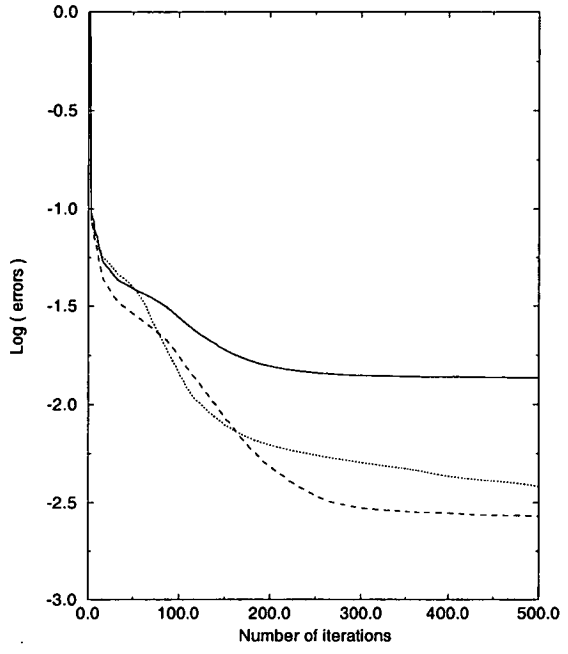


Figure 7.25: The performance of the conjugate gradient algorithm in the photon-limited case. The solid line, the dotted line and the dashed line represents 10^5, 5×10^5 *and* 10^6 *photon counts respectively.*

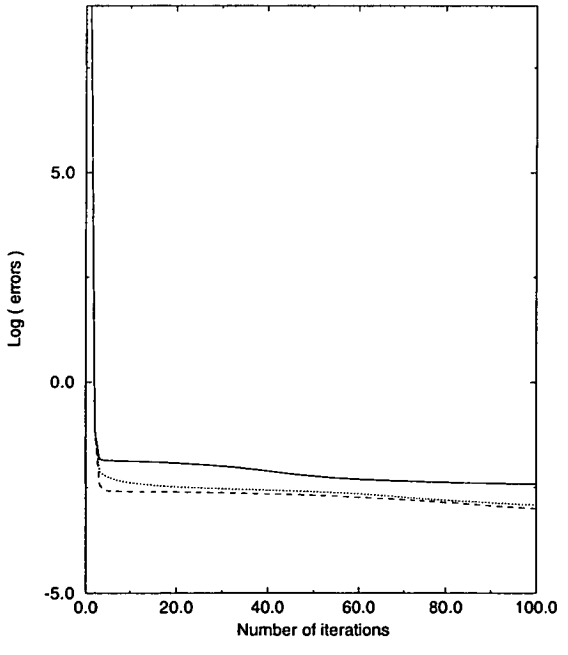


Figure 7.26: The performance of the projection-based algorithm in the photon-limited case. The solid line, the dotted line and the dashed line represents 10^5, 5×10^5 *and* 10^6 *photon counts respectively.*

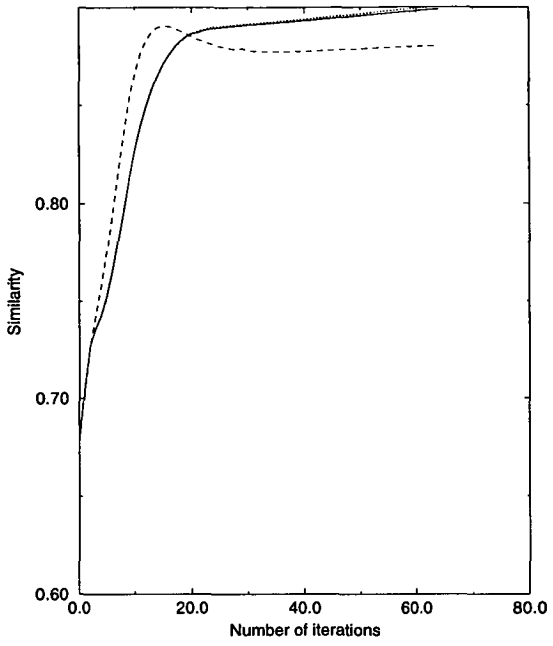


Figure 7.27: The convergence behaviour of the similarity metric for the maximum likelihood algorithm in the photon-limited case. The solid line, the dotted line and the dashed line represents 10^5, 5×10^5 *and* 10^6 *photon counts respectively.*

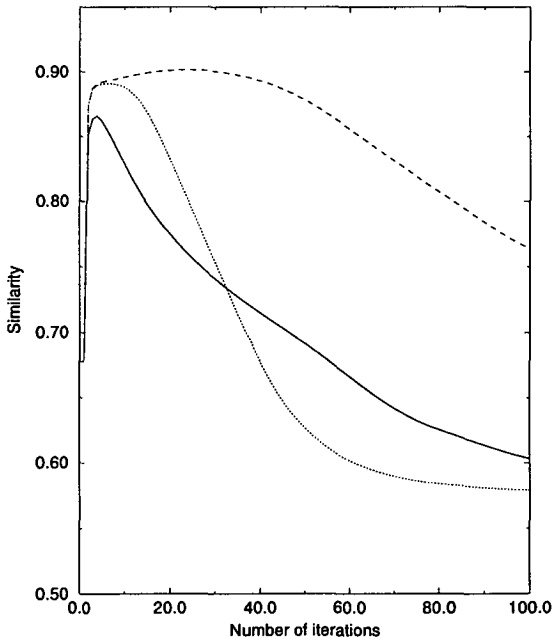


Figure 7.28: The convergence behaviour of the similarity metric for the projection-based algorithm in the photon-limited case. The solid line, the dotted line and the dashed line represents 10^5, 5×10^5 *and* 10^6 *photon counts respectively.*

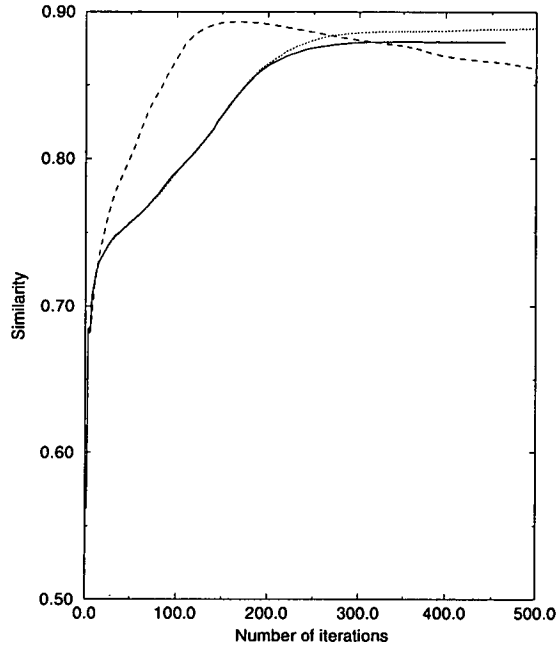


Figure 7.29: The convergence behaviour of the similarity metric for the conjugate gradient algorithm in the photon-limited case. The solid line, the dotted line and the dashed line represents 10^5, 5×10^5 *and* 10^6 *photon counts respectively.*

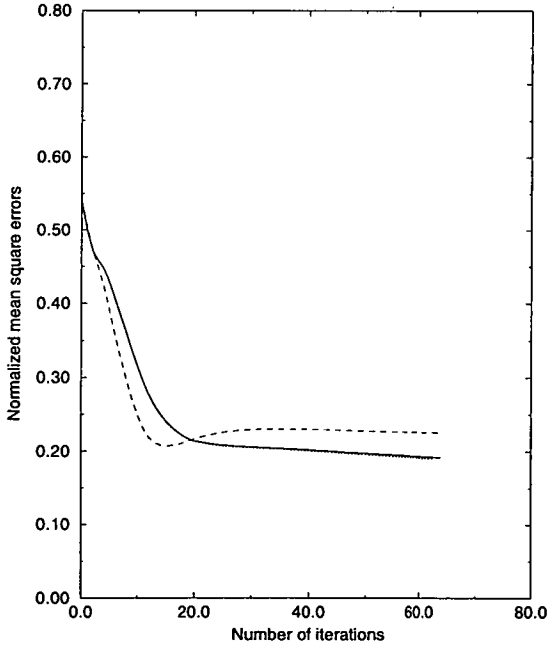


Figure 7.30: The convergence behaviour of the normalized mean square error for the maximum likelihood algorithm in the photon-limited case. The solid line, the dotted line and the dashed line represents 10^5, 5×10^5 *and* 10^6 *photon counts respectively.*

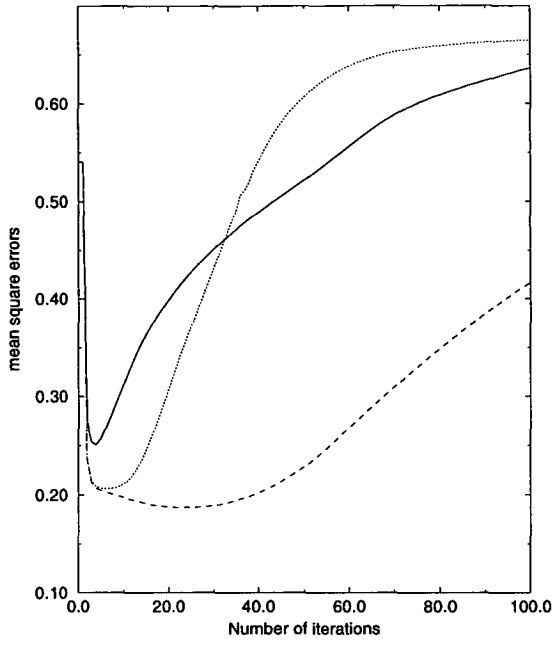


Figure 7.31: The convergence behaviour of the normalized mean square error for the projection-based algorithm in the photon-limited case. The solid line, the dotted line and the dashed line represents 10^5, 5×10^5 *and* 10^6 *photon counts respectively.*

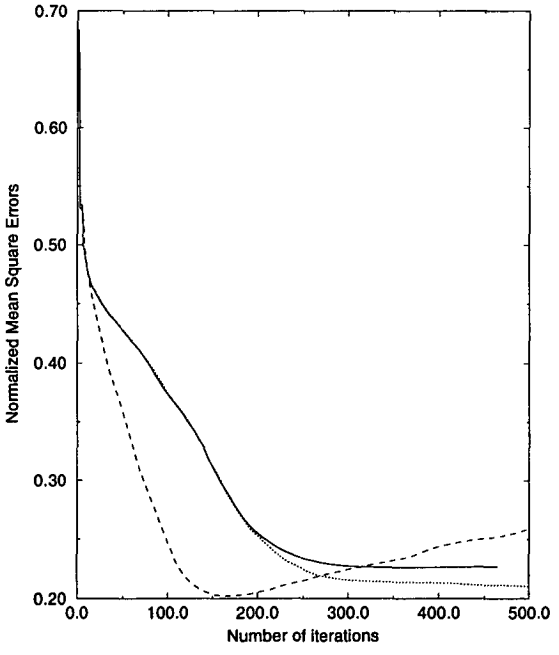


Figure 7.32: The convergence behaviour of the normalized mean square error for the conjugate gradient algorithm in the photon-limited case. The solid line, the dotted line and the dashed line represents 10^5, 5×10^5 *and* 10^6 *photon counts respectively.*

	Photon counts	Peak similarity	Number of iterations taken
Maximum Likelihood Algorithm	100000	0.8988	64
	500000	0.8999	63
	1000000	0.8999	16
Conjugate Gradient Algorithm	100000	0.8796	350
	500000	0.8887	500
	1000000	0.8930	170
Projection Based Algorithm	100000	0.8657	4
	500000	0.8909	6
	1000000	0.9015	23

Table 7.14: Peak similarity metrics for the three algorithms in the presence of different noise level.

would increase in the presence of noise. Similarly, the decreases in error function values in both the conjugate gradient algorithm and the projection-based algorithm does not guarantee that the similarity metric would increase. This effect is known as superresolution in which one tries to extrapolate the value of the object in the frequency domain beyond a particular frequency limit determined by noise. As noted by Sementilli et al [102], superresolution, even when the point spread function is known, is prone to a null image artifact if iterations are carried too far. This artifact presents itself as high frequency noise. According to Lannes et al [103], only a smooth version of the original object can be recovered in severe noise contamination situation as noise and the object in the high frequency region are indistinguishable. This artifact, however, can be avoided by simply not reducing the error to below the level predicted by the consideration of the noise [79]. This artifact is encountered by nearly all blind deconvolution algorithms in the presence of noise, with the exception of the algorithm employed the Wiener filtering technique [44,45] in which the use of a Wiener filter in the processing loop has the effect of suppressing this artifact.

Chapter 8

Conclusions And Suggestions For Further Research

This thesis has been concerned with astronomical imaging. In particular, two approaches, namely the real time adaptive optics approach and the blind deconvolution approach, have been investigated in Chapters 4-7. This chapter presents a summary of the major results that have been obtained so far. Furthermore, several avenues for continuing research are suggested.

8.1 Conclusions

The problem that originates from the transmission of light through the turbulent atmosphere is discussed in this thesis. Although the atmospheric turbulence affects both the amplitude and the phase, its main effect on the quality of the image is the random phase aberration. Many efforts have been made to compensate the phase distortion so that the qualtiy of the image obtained with ground based telescope can be improved. There are generally two approaches to do so. The first one is a real time processing approach, commonly known as adaptive optics approach. It has been discussed in Chapters 4-5. The second approach is the post-detection processing technique which is discussed in Chapters 6-7.

The adaptive optics approach is itself a complex technology. It ompensates the effect of the atmospheric turbulence through the use of hardware design, that is, by introducing disturbance that is calculated to be just the opposite of the atmospheric disturbance so that both disturbances cancel each other out and thus produce a good quality picture. Hence the estimation of the disturbance introduced by the atmospheric turbulence is an important step in this approach.

One conventional way to do so is to reformulate the estimation problem as a least square problem. A pseudo-inverse solution can then be found. However, this formulation makes the problem under-determined and thus the estimated wavefront is often not very accurate. In order to solve this under-determined problem, some extra constraints need to be applied so that the estimation problem becomes well-conditioned. One a priori information that is available is the covariance matrix of the weightings of the Zernike polynomials. It can be obtained theoretically by assuming Kolmogorov turbulence model. This leads the estimation to be over-determined and a good estimation of wavefront is resulted as is shown in Chapter 4.

The second approach is the post-detection processing technique. The basic principle is to freeze the atmospheric turbulence by obtaining a sequence of images using short exposure time. This method is founded upon the insight that high spatial frequency components are preserved in these short exposure images. An ensemble of these short exposure images could have all the spatial frequencies passed by the telescope aperture and could be used to recover the diffraction-limited object.

The conventional post-detection processing methods can be classified as a deconvolution problem in which the blurring function modelling the effect of the atmospheric turbulence is estimated separately by some means. Examples of these method include the speckle interferometry and wavefront sensing deconvolution. There are, however, another approach in the post-detection processing, namely the blind deconvolution approach. Blind deconvolution problem refers to a problem in which both the object of interest and the blurring function need to be estimated simultaneously from the recorded image. Three different approaches for solving the blind deconvolution problem are the conjugate gradient method, the maximum likelihood method using the GEM algorithm and the projection-based method. They are discussed in Chapter 6.

These three methods can be formulated using the Bayes theorem but with different

assumption made on the noise statistics and different application of the a priori information. The Gaussian noise is assumed in both the conjugate gradient and the projection-based methods, while the Poisson noise statistics is assumed in the maximum likelihood approach. In the application of the a priori information such as the positivity constraint, both the maximum likelihood and the projection based methods treat that absolutely while the conjugate gradient method treats that as an inexact information. How appropriate this assumption may be will depend on the real situation.

A comparison of these three methods is made in Chapter 7. The comparison is based on the quality of the reconstruction and the convergence rate of the algorithm. It is found that in general, all three methods can blindly deconvolve a blurred image and the quality of the reconstruction is improved if more number of frames are used in the algorithm. The noise performance of these algorithms is also investigated. It is found that all three algorithms experience the superresolution effect as a result of the low SNR in the high frequency region.

8.2 Suggestions For Further Research

In Chapter 6, all the three methods proposed to solve the blind deconvolution problems are iterative in nature. Several difficulties might encounter in practice. For example, the convergence rate is often slow. It becomes particularly ineffective when the size of the problem is large.

While the cause for slow convergence is complex, a possible reason is that the iterative procedures operate only on the fine grid of fine resolution, thus reducing only the high frequency but not the low frequency content of the estimation error. Hence one way to solve this difficulty is to apply multi-grid processing technique as proposed by Wang et. al [104].

Wavelet transform based techniques have been applied in many different areas, such as image coding and object recognition. One desirable feature of the wavelet transform is that it provides a multiresolution view of an object and thus enables multi-grid processing. By decomposing the original 'single' channel problem into a 'multi' channel problem, the restoration algorithm can also be easily adapted to the properties of the signal and the noise in each subband.

Also, as most of the signals in practice have their energy concentrated at the low frequency regions rather than the high frequency regions, a method of successively restoring/deconvolving the true image from the low frequency part to the high frequency part could be employed. This successive restoration/deconvolution of the true signal would provide a successive better approximation to the true signal at each level.

The Mallat pyramid algorithm [105] could be used to implement the orthonormal wavelet transform and thus decompose signals into multi-channel. The deconvolution problem could be formulated in the wavelet domain as,

$$Wg = WHW^{-1}Wf + Wc \tag{8.1}$$

where W denotes the wavelet operator. The term WHW^{-1} can be obtained by first applying the wavelet transform to each row of H and then to each column of the result. It can, however, be obtained efficiently in Fourier domain as it has the semi-block circular structure [106].

I plan to explore the use of wavelet transform techniques in solving the blind deconvolution problem. In particular, by transforming the problem into the wavelet domain as shown in Eq (8.1), it becomes a multi-grid processing and the size of the problem is reduced which should improve the convergence rate.

References

[1] H. Baher, *Analog And Digital Signal Processing.* John Wiley And Sons, 1990.

[2] R. N. Bracewell, *The Fourier Transform and Its Applications.* McGraw-Hill International Editions, 1986.

[3] A. Oppenheim and J. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, pp. 529–541, May 1981.

[4] J. W. Goodman, *Introduction to Fourier Optics*, vol. xiii, 287 of *McGraw-Hill physical and quantum electronics.* McGraw-Hill Book Company, 1968.

[5] J. W. Goodman, *Statistical Optics.* United States of America: John Wiley and Sons, 1985.

[6] J. Mariotti, "Introduction to fourier optics and coherence," in *Diffraction-Limited Imaging with Very Large Telescopes* (D. Alloin and J. Mariotti, eds.), (F-921295 Meudon Principal France), pp. 3–31, Kluwer Academic Publishers, 1988.

[7] R. J. Noll, "Zernike polynomials and atmospheric turbulence," *Journal of the Optical Society of America*, vol. 66, pp. 207–211, Mar. 1976.

[8] P. McCullagh and J. Nelder, *Generalized Linear Models.* Chapman and Hall, 2 ed., 1991.

[9] G. B. Wetherill, *Intermediate Statistical Methods.* Chapman and Hall Ltd, 1981.

[10] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of Royal Statistics Society B*, vol. 39, pp. 1–38, 1977.

[11] C. J. Wu, "On the convergence properties of the em algorithm," *Annual Statistics*, vol. 11, pp. 95–103, 1983.

[12] R. Lagendijk, J. Biemond, and D. E. Boekee, "Identification and restoration of noisy blurred images using em algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 1180–1191, July 1990.

[13] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 1, pp. 113–122, Oct. 1982.

[14] T. J. Hebert and K. Lu, "Expectation-maximization algorithms, null spaces, and map image restoration," *IEEE Transactions on Image Processing*, vol. 4, pp. 1984–1095, Aug. 1995.

[15] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, pp. 2024–2036, Dec. 1989.

[16] T. Hebert and R. Leahy, "A generalized em algorithm for 3-d bayesian reconstruction from poisson data using gibbs priors," *IEEE Transactions on Medical Imaging*, vol. 8, pp. 194–202, June 1989.

[17] P. Gill, W. Murray, and M. Wright, *Practical Optimization.* Academic Press, 1984.

[18] B. Gottfried and J. Weisman, *Introduction to Optimization Theory.* Englewood Cliff: Prentice-Hall Inc, 1973.

[19] L. Hageman and D. Young, *Applied Iterative Methods.* Academic Press, 1981.

[20] P. Gill and W. Murray, *Numerical Methods for Constrained Optimization.* Academic Press, 1974.

[21] R. Fletcher, *Practical Methods of Optimization.* John Wiley and Sons, 1987.

[22] M. Sezan and H. Stark, "Image restoration by the method of convex projections : Part 2 - applications and numerical results," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 95–101, 1982.

[23] A. Levi and H. Stark, "Image restoration by the method of generalized projections with application to restoration from magnitude," *Journal of the Optical Society of America A*, vol. 1, pp. 932–943, Sept. 1984.

[24] H. Trussell and R. Civanlar, "The feasible solution in signal restoration," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 32, pp. 201–212, April 1984.

[25] R. Lagendijk, J. Biemond, and D. E. Boekee, "Regularized iterative image restoration with ringing reduction," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 36, pp. 1874–2887, Dec. 1988.

[26] J. Biemond, R. L. Lagendijk, and R. M. Mersereau, "Iterative methods for image deblurring," *Proceedings of the IEEE*, vol. 78, pp. 856–883, May 1990.

[27] D. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1 - theory," *IEEE Transactions on Medical Imaging*, vol. MI-1, pp. 81–94, Oct. 1982.

[28] H. Peng and H. Stark, "Signal recovery with similarity constraints," *Journal of the Optical Society of America A*, vol. 6, pp. 844–851, June 1989.

[29] R. K. Tyson, *Principles of Adaptive Optics.* 1250 Sixth Avenue, San Diego, CA 92101: Academic Press, INC., 1991.

[30] F. Roddier, "The effect of atmospheric turbulence in optical astronomy," in *Progress in Optics XIX* (E. Wolf, ed.), pp. 282–376, North-Holland, Amsterdam, 1981.

[31] R. Q. Fugate, "Laser beacon adaptive optics," *Opitcs and Photonics News*, pp. 14–19, June 1993.

[32] R. R. Parenti, "Adaptive optics for astronomy," *The Lincoln Laboratory Journal*, vol. 5, no. 1, pp. 93–113, 1992.

[33] F. Merkle, "Adaptive optics," *Physics World*, pp. 33–38, Jan. 1991.

[34] T. Lawrence, D. Goodman, E. Johansson, and J. Fitch, "Speckle imaging of satellites at the us air force maui optical station," *Applied Optics*, vol. 31, no. 29, pp. 6307–6321, 1992.

[35] A. Labeyrie, "Attainment of diffraction limited resolution in large telescopes by fourier analysing speckle patterns in star images," *Astronomy and Astrophysics*, vol. 6, pp. 85–87, 1970.

[36] C. Schwartz, G. Baum, and E. Ribak, "Turbulence-degraded wave fronts as fractal surfaces," *Journal of the Optical Society of America A*, vol. 11, pp. 444–451, Jan. 1994.

[37] R. Lane, A. Glindemann, and J. Dainty, "Simulation of a kolmogorov phase screen," *Waves in Random Media*, vol. 2, pp. 209–224, 1992.

[38] R. R. Parenti and R. J. Sasiela, "Laser guide star systems for astronomical applications," *Journal of the Optical Society of America A*, vol. 11, pp. 288–309, Jan. 1994.

[39] D. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," *Journal of the Optical Society of America*, vol. 56, pp. 1372–1379, Oct. 1966.

[40] R. Bates and M. McDonnell, *Image Restoration and Reconstruction*, vol. 16 of *The Oxford Engineering Science Series.* Clarendon Press Oxford, 1989.

[41] Y. Bruck and L. Sodin, "On the ambiguity of the image reconstruction problem," *Optics Communications*, vol. 30, pp. 304–308, Sept. 1979.

[42] M. Hayes and J. McCLellan, "Reducible polynomials in more than one variable," *Proceedings of the IEEE*, vol. 70, no. 2, pp. 197–198, 1982.

[43] A. Huiser and P. V. Toorn, "Ambiguity of the phase reconstruction problem," *Optics Letters*, vol. 5, pp. 499–501, Nov. 1980.

[44] G. Ayers and J. Dainty, "An iterative blind deconvolution method and its applications," *Optics Letter*, vol. 13, pp. 547–549, 1988.

[45] B. Davey, R. Lane, and R. Bates, "Blind deconvolution of noisy complex-valued image," *Optics Communications*, vol. 69, pp. 353–356, Jan. 1989.

[46] R. Lane and R. Bates, "Automatic multidimensional deconvolution," *Journal of the Optical Society of America A*, vol. 4, pp. 180–188, Jan. 1987.

[47] H. Babcock, "The possibility of compensating astronomical seeing," *Publications of the Astronomical Society of the Pacific*, vol. 65, pp. 229–236, Oct. 1953.

[48] R. Foy and A. Labeyrie, "Feasibility of adaptive telescope with laser probe," *Astronomy and Astrophysics*, vol. 152, pp. 29–31, 1985.

[49] J. Hardy, "Active optics: A new technology for the control of light," *Proceedings of he IEEE*, vol. 66, pp. 651–697, June 1978.

[50] F. J. Dryson, "Photon noise and atmospheric noise in active optical systems," *Journal of the Optical Society of America*, vol. 65, pp. 551–558, May 1975.

[51] J. Primot, G. Rousset, and J. Fontanella, "Deconvolution from wavefront sensing: a new technique for compensating turbulence-degraded images," *Journal of the Optical Society of America A*, vol. 7, pp. 1598–1608, Sept. 1990.

[52] R. Hudgin, "Wavefront reconstruction for compensated imaging," *Journal of the Optical Society of America*, vol. 67, pp. 375–378, 1977.

[53] D. Fried, "Least square fitting of a wavefront distortion estimate to an array of phase difference measurements," *Journal of the Optical Society of America*, vol. 67, pp. 370–375, Mar. 1977.

[54] R. J. Noll, "Phase estimates from slope type wavefront sensors," *Journal of the Optical Society of America*, vol. 68, pp. 139–140, 1978.

[55] B. Hunt, "Matrix formulation of the reconstruction of phase values from phase differences," *Journal of the Optical Society of America*, vol. 69, pp. 393–399, 1979.

[56] W. Southwell, "Wavefront estimation from wavefront slope measurements," *Journal of the Optical Society of America*, vol. 70, pp. 998–1006, Aug. 1980.

[57] M. Ealey and J. Washeba, "Continuous facesheet low voltage deformable mirrors," *Optical Engineering*, vol. 29, pp. 1191–1198, Oct. 1990.

[58] J. M. Beckers, "Adaptive optics for astronomy: Principles, performance and applications," *Annual Reviews of Astronomy and Astrophysics*, vol. 31, 1993.

[59] F. Merkle, "Real time wavefront sensing and adaptive optics," in *Diffraction-Limited Imaging with Very Large Telescopes* (D. Alloin and J. Mariotti, eds.), (Karl-Schwarzschild-Str.2, K-8046 Garching, Federal Republic of Germany), pp. 237–248, Kluwer Academic Publishers, 1988.

[60] C. Perrier, "Amplitude estimation from speckle interferometry," in *Diffraction-Limited Imaging with Very Large Telescopes* (D. Alloin and J. Mariotti, eds.), (Av. Charles Andre, F-69561 Saint-Genis-Laval Cedex France), pp. 99–111, Kluwer Academic Publishers, 1988.

[61] R. Gerchberg and W. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 2, pp. 237–246, 1972.

[62] J. Fienup, "Space object imaging through the turbulent atmosphere," *Optical Engineering*, vol. 18, no. 5, pp. 529–534, 1979.

[63] J. Fienup, "Phase retrieval algorithms : A comparison," *Applied Optics*, vol. 21, pp. 2758–2769, Aug. 1982.

[64] K. Knox and B. Thompson, "Recovery of images from atmospherically degraded short exposure photographs," *Astrophysical Journal*, vol. 193, pp. L45–L48, Oct. 1974.

[65] G. Ayers, M. Northcott, and J. Dainty, "Knox-thompson and triple correlation imaging through atmospheric turbulence," *Journal of the Optical Society of America A*, vol. 5, pp. 963–985, July 1988.

[66] P. Nisenson, "Speckle imaging with the papa detector and the knox-thompson algorithm," in *Diffraction-Limited Imaging with Very Large Telescopes* (D. Alloin and J. Mariotti, eds.), (60 Garden St. Cambridge MA 02138 U.S.A.), pp. 157–169, Kluwer Academic Publishers, 1988.

[67] A. Chelli, "Merits of phase restoration methods," in *Diffraction-Limited Imaging with Very Large Telescopes* (D. Alloin and J. Mariotti, eds.), (Apartado Postal 70-264 Mexico D.F. 04510 Mexcio), pp. 201–220, Kluwer Academic Publishers, 1988.

[68] A. W. Lohmann, G. Weigelt, and B. Wimitzer, "Speckle masking in astronomy: Triple correlation theory and applications," *Applied Optics*, vol. 22, pp. 4028–4037, Dec. 1983.

[69] A. W. Lohmann and B. Wirnitzer, "Triple correlations," *Proceedings of the IEEE*, vol. 72, pp. 899–901, July 1984.

[70] K. Hofmann and G. Weigelt, "Iterative image reconstruction from the bispectrum," *Astronomy and Astrophysics*, vol. 278, pp. 328–339, 1993.

[71] M. Northcott, G. Ayers, and J. Dainty, "Algorithms for image reconstruction from photon-limited data using the triple correlation," *Journal of the Optical Society of America A*, vol. 5, pp. 986–992, July 1988.

[72] G. Weigelt, "Speckle masking, speckle spectroscopy, and optical aperture synthesis," in *Diffraction-Limited Imaging with Very Large Telescopes* (D. Alloin and J. Mariotti, eds.), (Erwin-Rommel-Str. 1 D-8520 Erlangen Fed. Rep of Germany), pp. 191–200, Kluwer Academic Publishers, 1988.

[73] J. Meng and G. J. Aitken, "Triple-correlation subplane reconstruction of photon address stellar images," *Journal of the Optical Society of America A*, vol. 7, pp. 1243–1250, July 1990.

[74] C. A. Haniff, "Least square fourier phase estimation from the modulus 2 π bispectrum phase," *Journal of the Optical Society of America A*, vol. 8, pp. 134–140, Jan. 1991.

[75] T. G. Stockham, T. M. Cannon, and R. B. Ingebretsen, "Blind deconvolution through digital signal processing," *Proceedings of the IEEE*, vol. 64, pp. 678–692, April 1975.

[76] R. Lane, *Blind Deconvolution and Phase Retrieval.* Christchurch, New Zealand: PhD Thesis, University of Canterbury,, 1988.

[77] J. Seldin and J. Fienup, "Iterative blind deconvolution algorithm applied to phase retrieval," *Journal of the Optical Society of America A*, vol. 7, pp. 428–433, Mar. 1990.

[78] R. Lane, "Blind deconvolution of speckle images," *Journal of the Optical Society of America A*, vol. 9, pp. 1508–1514, 1992.

[79] N. Law and R. Lane, "Blind deconvolution using least squares minimisation," *Optics Communications*, vol. 128, pp. 341–352, 1996.

[80] E. Wallner, "Optimal wavefront correction using slope measurements," *Journal of the Optical Society of America*, vol. 73, pp. 1771–1776, 1983.

[81] C. Solomon, N. Wooder, and J. Dainty, "Bayesian estimation of atmospherically distorted wavefronts," *Opitcal Review*, vol. 2, no. 3, 1995.

[82] R. Lane, N. Law, and A. Bainbridge-Smith, "Ensemble deconvolution using a wavefront sensor," in *Proceedings of DICTA-93* (K. K. Fung and A. Ginige, eds.), (GPO Box 252C, Hobart, Tasmania 7001, Australia), pp. 236–243, Australian Pattern Recognition Society, Dec. 1993.

[83] D. Fried, "Statistics of a geometric representation of wavefront distortion," *Journal of the Optical Society of America*, vol. 55, pp. 1427–1435, 1965.

[84] B. Frieden, *Probability, Statistical Optics and Data Testing.* Springer-Verlag, 1991.

[85] J. Wang and J. Markay, "Modal compensation of atmospheric turbulence phase distortion," *Journal of the Optical Society of America*, vol. 68, pp. 78–87, Jan. 1978.

[86] B. M. Welsh and C. S. Gardner, "Performance analysis of adaptive optics systems using laser guide stars and slope sensors," *Journal of the Optical Society of America A*, vol. 6, pp. 1913–1923, Dec. 1989.

[87] H. Sorenson, *Parameter Estimation.* New York: Marcel Dekker, 1980.

[88] E. Thiebaut and J. Conan, "Strict a priori constraints for maximum likelihood blind deconvolution," *Journal of the Optical Society of America A*, vol. 12, pp. 485–492, Mar. 1995.

[89] S. Jefferies and J. Christou, "Restoration of astronomical images by iterative blind deconvolution," *Astrophysical Journal*, vol. 415, pp. 862–874, 1993.

[90] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge Univeristy Press, 1992.

[91] Y. Yang, N. Galatsanos, and H. Stark, "Projection-based blind deconvolution," *Journal of the Optical Society of America A*, vol. 11, pp. 2401–2409, Sept. 1994.

[92] T. Schulz, "Multiframe blind deconvolution of astronomical images," *Journal of the Optical Society of America A*, vol. 10, pp. 1064–1077, May 1993.

[93] N. Law and D. Nguyen, "Multiple frame projection based blind deconvolution," *Electronics Letter*, vol. 31, pp. 1732–1733, Sept. 1995.

[94] R. Lagendijk, R. Mersereau, and J. Biemond, "On increasing the convergence rate of regularized iterative image restoration algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 1183–1186, 1987.

[95] Y.-H. Pao, *Adaptive Pattern Recognition and Neural Networks.* Addison Wesley Publishing Company, Inc., 1989.

[96] J. Zurada, *Introduction to Artifical Neural Systems.* USA: West Publishing Company, 1992.

[97] D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing Explorations in the Microstructure of Cognition Volume 1 Foundations.* USA: Massachusetts Institute of Technology, 1986.

[98] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Networks*, vol. 1, pp. 295–307, 1988.

[99] N. Law and D. Nguyen, "Improved convergence of the projection based blind deconvolution," *Electronics Letter*, vol. 31, pp. 1734–1735, Sept. 1995.

[100] N. Miura, S. Kuwamura, N. Baba, S. Isobe, and M. Noguchi, "Parallel scheme of the iterative blind deconvolution method for stellar object reconstruction," *Applied Optics*, vol. 32, pp. 6514–6520, Nov. 1993.

[101] A. L. N.F. Law and D. Nguyen, "Multiresolution deconvolution using conjugate gradient," *to be published in IEEE Singapore International Conference on Communication Systems*, 1996.

[102] P. Sementilli, B. Hunt, and M. Nadar, "Analysis of the limit to superresolution in incoherent imaging," *Journal of the Optical Society of America A*, vol. 10, pp. 2265–2276, 1993.

[103] A. Lannes, S. Roques, and M. Casanove, "Resolution and robustness in image processing : A new regularization principle," *Journal of Optical Society America*, vol. 4, pp. 189–199, 1987.

[104] G. Wang, J. Zhang, and G. W. Pan, "Solution of inverse problems in image processing by wavelet expansion," *IEEE Transactions on Image Processing*, vol. 4, pp. 579 – 593, May 1995.

[105] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, July 1989.

[106] M. R. Banham, N. P. Galatsanos, H. L. Gonzalez, and A. K. Katsaggelos, "Multichannel restoration of single channel images using a wavelet-based subband decomposition," *IEEE Transactions on Image Processing*, vol. 3, pp. 821–833, Nov. 1994.